UNIVERSITY OF THE WEST OF ENGLAND

ROBOT MEDIATED COMMUNICATION: ENHANCING TELE-PRESENCE USING AN AVATAR

HAMZAH ZIYAAD HOSSEN MAMODE

10/08/2015

A thesis submitted in partial fulfilment of the requirements of the University of the West of England, Bristol for the degree of Doctor of Philosophy

Faculty of Environment and Technology

University of the West of England, Bristol

August 2015

“If words of command are not clear and distinct, if orders are not thoroughly understood, then the general is to blame.”

– Sun Tzu (c. 6th century BCE)

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Signed: --

Date: ---

Abstract

In the past few years there has been a lot of development in the field of tele-presence. These developments have caused tele-presence technologies to become easily accessible and also for the experience to be enhanced. Since tele-presence is not only used for tele-presence assisted group meetings but also in some forms of Computer Supported Cooperative Work (CSCW), these activities have also been facilitated. One of the lingering issues has to do with how to properly transmit presence of non-co-located members to the rest of the group. Using current commercially available tele-presence technology it is possible to exhibit a limited level of social presence but no physical presence. In order to cater for this lack of presence a system is implemented here using tele-operated robots as avatars for remote team members and had its efficacy tested. This testing includes both the level of presence that can be exhibited by robot avatars but also how the efficacy of these robots for this task changes depending on the morphology of the robot. Using different types of robots, a humanoid robot and an industrial robot arm, as tele-presence avatars, it is found that the humanoid robot using an appropriate control system is better at exhibiting a social presence. Further, when compared to a voice only scenario, both robots proved significantly better than with only voice in terms of both cooperative task solving and social presence. These results indicate that using an appropriate control system, a humanoid robot can be better than an industrial robot in these types of tasks and the validity of aiming for a humanoid design behaving in a human-like way in order to emulate social interactions that are closer to human norms. This has implications for the design of autonomous socially interactive robot systems.

Acknowledgements

I would like to thank Dr. Paul Bremner, Prof. Anthony Graham Pipe and Dr. Brian Carse, from the University of the West of England, and Dr. Mike Fraser and Dr. Sriram Subramanian, from the University of Bristol for their invaluable support and guidance over the years.

I would also like to thank my parents and sister for their unconditional moral support.

Table of Contents

Figures

Tables

Glossary

Avatar	A representation of a being in an environment that the being is not physically present in.
CSCW	Computer Supported Cooperative Work, fully described in section 2.2.5.
Embodiment	A physical representation of the actions of a being in an environment.
HCI	Human-Computer Interaction, the study of how people interact with computers and how computers can be developed to allow successful interaction.
HHI	Human-Human Interaction, the study of how people interact with each other and how this interaction can be affected by the setting in which it happens.
HRI	Human-Robot Interaction, fully described in section 2.2.1.
HTI	Human-Tabletop Interaction, a subset of HCI focussed on how humans interact with Tabletop Interactive Devices.
Robot	A mechanical, electronic, or virtual artificial entity that is designed to accomplish a specific purpose.
SDK	Software Development Kit, a piece of software that helps in the development of other software that will use a specific system resource.
TID	Tabletop Interactive Device, fully described in section 3.3.

Video Conferencing	An ensemble of technologies used to capture and reproduce image from one area to another in order to allow groups of people to interact using an audio-visual representation of each other.
VoIP	Voice over Internet Protocol, a group of technologies used to enable the delivery of audio communication in between places connected using an internet based communication medium.

Chapter 1 Introduction

1.1 Chapter Introduction

Since its commercial debut in the early 1970s, videoconferencing has come a long way (Egido, 1988). While in the past such technology was only available to a select few that could afford the expensive equipment it required, videoconferencing is now available to anyone that has access to a suitable computer or smart phone. During this time the reason for its use, that is, being able to also see the interlocutor instead of just hearing them, has remained the same. Advances in technology has allowed the equipment required for such an activity to be reduced from a setup that required half a room, to a piece of equipment that can be held in one's hand (Judge & Neustaedter, 2010).

These advances in technology are not limited to just miniaturisation, novel methods have been developed to allow geographically distributed groups of people to interact together. These novel methods have a common theme of allowing the individuals to share a common space; however, the way through which the sharing is enabled can be quite different. One of these ways is through the use of virtual environments, in which and with which the individuals can interact. While this type of interaction can be quite common, such as with online games (Brown & Bell, 2004), more immersive methods have also been implemented (Hindmarsh, Fraser, Heath, Benford, & Greenhalgh, 1998). While these virtual environments support interaction either with the environment itself or with other individuals represented in them, through the use of avatars, they nevertheless allow the user to have a tele-presence experience (Witmer & Singer, 1998). This interaction through a tele-presence experience can be quite different from the real world interaction since the virtual environment may not have the same properties as a real world medium (Gaver, 1992). Consequently, interacting in virtual environments may require training of the interlocutors and an increase in cognitive load that may impact the interaction style.

Since real world environments are what individuals usually used to interact with each other, technologies have been developed to enable the tele-presence experience in the real world. In the last decade, new advances in technology has allowed for the development of services such as Musion Tele-presence from Musion[1], which makes use of holographic technology. Using such technology allows the interlocutors to almost share a common physical space. However, such setups which allow a 3D holographic rendition of remote interlocutors is highly restricted due to both the price of the hardware requirements and also the presence of support technology in the region it is deployed, such as, high bandwidth connections. Perhaps most importantly, since the holographic representation is only visual, not physical, object manipulation and other interactions only enabled via 3D physical presence, are not possible.

An alternative approach to enable geo-distributed groups to interact was made famous by Professor Hiroshi Ishiguro (Ishiguro Hiroshi) of Osaka University where humanoid robots are used as real world avatars (Guizzo, 2010). In this situation, the professor uses a life size humanoid robot, with facial features similar to his, to give lectures in a room where the robot avatar is present, while he is physically present in a different location (Guizzo, 2010). While the robot uses gestures to enhance the presence on behalf of the professor, audio communication is also enabled between the different locations (Guizzo, 2010). This example of real world tele-presence using robots, while adequate for the task, also provides a peek into potential issues of using such a setup. As mentioned earlier, the facial features of the robot is very human like, consequently one major issue with this robot is the negative acceptance of the robot due to the uncanny valley (More information provided in section 2.2.3), whereby objects that may look too human like while not being human can be rejected by an observer (Mori, 1970). Moreover, the use of such a robot, intended to be a

[1] Musion is a product of Vimeo that uses holograms to portray remote users. ***www.musion.co.uk***

simulacrum, could be quite confusing if the robot is used by another person, as the representation may not match the operator.

Despite, these possible issues, the use of robots as local physical representation of remote individuals may provide some definite unique advantages. Before delving into the advantages of the use of robots in such situations, a brief description of the features that robots that make them suitable for such uses is provided. Robots, as they are used in this thesis, can be defined as physical mechanical entities that can be programmed to perform specific tasks. Since robots are assembled, they can be designed for a specific function in mind and consequently, there are very many different types of robots, both in terms of form but also in terms of size. However, independent of these factors, a robot usually runs control programs that allow it to perform specific actions following triggers.

Though these control programs have been used since the early 1950s, they have also been used in recent years to make such robots seem intelligent by allowing them to respond autonomously instead of being controlled. These program-controlled robots have been used, apart from simulating intelligent behaviour, to portray a level of presence, more specifically social presence, allowing people to interact with them in a limited fashion. This same type of implementation has also been used with a human being controlling the robots' responses instead of an automated program, usually requiring the human operator to take an active part in the control of the robot. This explicit control mechanism uses specific equipment requiring the operator to manually control the responses of the robot. However, if a robot is to adequately represent the presence of an individual, a more intuitive control will be required. Using an implicit control system, the movements of a human being will be able to be captured, by such a system, and portrayed by a humanoid robot. The advantages of such a system lie in having a real world presence, through the robot, while at the same time allowing the person to behave as they would usually do and not having to think how their actions need to be represented on such a robot.

In recent years, there has been an exponential growth in the use of robots in general but even more so in the sector of service robots and according to the United Nations Economic Commission (UNEC) and International Federation of Robotics (IFR), this trend is going to persist at least until 2025 (United Nations Economic Commission for Europe;The International Federation of Robotics, 2002). While some of these service robots are merely domestic robots such as automated cleaning robots, a large portion of them are currently toys and entertainment robots. In future years, it is expected that assistive robots will become more commonplace (Bartneck & Forlizzi, 2004). While not all of these robots are going to be humanoid robots, they will probably be composed of a high number of degrees of freedom to be able to perform their tasks adequately. One alternative purpose that these robots could provide, is to act as physical embodiments in the context of remote interaction. With this in mind, it is important to understand how the impact of the use of human inspired behaviours for humanoid robots differ from the default behaviour of non-humanoid robots in the context of robotic physical embodiments for remote social tele-presence.

1.2 Tele-presence, Robotics and Tabletop Devices

The concept of tele-presence has been around for at least half a century, implemented in one form or another. What started with the mere ability to hear voice of a person, who is not in the immediate vicinity, has developed into technologies that now allows people to see each other when communicating over vast distances. Using these technologies, it is possible to see, in real-time, what other people from around the world are doing and interact with them to a certain extent. With current video conferencing systems one can see the remote interlocutors and this allows for a higher level of social presence than with voice only communication (Heath & Luff, 1991). Although the representation is limited to a screen, a level of trust can be established that is higher than with voice only communication, since tells, indications of someone being deceptive, can be transmitted using video support (Bos, Olson, Gergle, Olson, & Wright, 2002). While video conferencing is not ideal for properly enabling social presence (Heath & Luff, 1991), it allows for the different participants to see each other's faces, an important element of human interaction (Kraut, Fussell, & Siegel, 2003).

For all its benefits over voice only communication, video conferencing however suffers from one major drawback, it may not transfer all the gestural information between the different locations (Lincoln, et al., 2009). While video conferencing supports gestures, through the visibility of gestures on the screen, some information pertaining to the correct comprehension of gestures may be corrupted or lost (Lincoln, et al., 2009). Moreover, as was pointed out by Cohen and Harrison, the lack of appropriate representation of gestures may lead the interlocutor to stop using them (Cohen & Harrison, 1973). One class of gestures, deictic gestures, either when being performed by the hands or with the gaze, is completely lost through video conferencing (Nguyen & Canny, 2007). Additionally, some other gestures are also severely affected by video conferencing such as co-verbal gestures, which will sometimes be intentionally not performed to limit the confusion, leading to reduced social engagement (Bos, Olson, Gergle, Olson, & Wright, 2002). This results in poor engagement and speech overlaps.

In Computer Supported Cooperative Work (CSCW), video conferencing technologies have been used together with tabletop interactive devices (TIDs) in order to test whether video conferencing setups provide net benefits with regard to cooperative task solving (Tuddenham & Robinson, 2007). While these setups provide an increase in cooperation over voice only, there persists a presence disparity between the remote members and the co-located members (Tang, Neustaedter, & Greenberg, 2006). This disparity leads to the remote members to be ostracised by the co-located members, leading to the in-group/out-group mentality (Bos, Shami, Olson, Cheshin, & Nan, 2004). This effect was attributed to the non-physicality of the presence being exhibited by the remote members in the co-located environment (Tang, Neustaedter, & Greenberg, 2006). While this effect was observed, no successful solution has so far been put forward to help attenuate the impact.

In order to remedy this problem a number of solutions using physical and quasi-physical (through the use of holographic technology) presence have been investigated. One such solution, and the principle focus of this thesis is the use of robotic platforms that are tele-operated by a remote person, which also acts as a possible solution to the problem of using co-verbal gestures with video conferencing systems. While this solution is by no means ideal, it provides the immediate benefit of providing a physical presence. These robots are able to replicate the gestures of a remote participant, thus demonstrating both a physical presence due to their nature and a social presence due to their task. These robots in conjunction with a TID, such as, the Microsoft Surface (See Section 3.3.1), the actual TID used in this thesis, is the solution proposed here to better facilitate co-operation between geo-separated teams. However, it suffers from certain limitations, which include limitations in the quality of reproduction of gestures, and also involve high setup costs. More on these issues will be elaborated in Chapter 2. Later on in this thesis, the efficacy of such systems for this purpose and the value of human inspired behaviours in robots are investigated.

1.3 Aim and Objectives

1.3.1 Motivation

In the previous section, it was stated that one option for tele-presence is the use of robots as tele-presence platforms for the embodiment of the actions of remote individuals. In the investigation described in this thesis different robots are going to be used as platforms for participants taking part in cooperative tasks as a group. It does so in order to try to address some of the issues currently observed in the literature relating to social presence and social interaction with remote interlocutors such as in Tang, Neustaedter, and Greenberg (2006).

When using classic video conferencing setups with the remote participants being represented by images on a screen, a number of issues are raised with respect to the location and orientation of the screen to mirror the physicality aspect of the presence as was discussed by Tang, Boyle, and Greenberg in their work using Mixed Presence Groupware (Tang, Boyle, & Greenberg, 2004). The solution they proposed to ameliorate the disparities in visibility and presence, using projected 2D representations of the remote participants, proved moderately successful at establishing a high level of social presence (Tang, Neustaedter, & Greenberg, 2006). This was attributed to the non-physical location occupied by the virtual bodies and arms. A logical solution to that problem is to make use of a robot that would be able to fill in the physical location problem and further establish a presence that the local participants can engage with. This inherent physical presence that robots display, i.e., a robot exhibits a physical presence just by being there, is important for further social presence (Adalgeirsson & Breazeal, 2010). There are many facets to the robot presence and will be further elaborated in Chapter 7.

The use of robots in a mediated communication setting is in no way a cheap endeavour, advancement in technology and the general use of robots is likely to make this type of interaction more easily available in the future. Moreover, the fact that the types of robots that were used as part of the investigation are multi-purposed robots and are more likely to become common place objects in the future, make them a feasible option for these kind of interactions. Hence, gaining understanding of the efficacy of different robot platforms as avatars will likely be of high utility going forward.

On a side note, the use of robots, in the vicinity of the task, instead of directly captured and reproduced images, is likely to be more useful since events on screen usually go unnoticed when they are in the periphery of vision (Gaver, 1992). This would allow the participants to focus more on the task at hand while still being able to get the gesture information being conveyed by the robots (Kraut, Fussell, & Siegel, 2003).

1.3.2 Research Question

Most investigations can usually be summarised to one research question that acts as a core around which the rest of the investigation is conducted, therefore with respect to this investigation, the research question is:

How does a robot embodiment, which is executing the actions from a remote individual, impact the overall interaction when performing a group activity?

1.3.3 Aim

Following from the research question the overall aim of the investigation is to investigate what impact does the presence of remotely operated robots, acting as embodiments of the tele-operator, have on group based activities.

1.3.4 Objectives

In order to accomplish this aim, a number of sub-questions to the research question has been identified and providing an answer to each of these questions act as de facto objectives of the investigation. These questions are:

- Q1: How do people perceive pointing from a robot compared to a human?
- Q2: How would a robot working as an avatar for a remote person be useful for group interaction?
- Q3: What impact does the nature and behaviour of a robot have on group interaction?
- Q4: What impact does the behaviour of a robot have on the perception of the embodied user?

These questions will be further elaborated in light of the findings in the related fields of study and act as precursors to the hypotheses that will be stated in Chapter 2. It is to be noted that these questions, while being of particular importance for the current investigation, are also of much interest in the rest of the research community as they provide answers that can be applied to other uses in robotics or related fields of research; this will be shown in Chapter 2.

1.4 Context and Scope

In analysing the aim and objectives of the investigation, it quickly becomes apparent that a proper context needs to be established in order to properly frame the investigation. The idea behind robot mediated communication is to use robots as embodiments, both as placeholders for the tele-operator of the robot and as a local actor to portray the actions of the tele-operator.

As a first step toward that goal it is important to define what types of robots will be used for this purpose. As stated previously, there are a wide variety of robots ranging from massive industrial robots to toy robots, it is therefore important to identify what types of robots will be most useful for the investigation. The first robot of these robots is a humanoid life sized robot, this robot will have the most human-like physical features and will therefore be a good candidate for a direct mapping between the actions of the robot to that of the tele-operator. While the advantages of using a humanoid robot will be further elaborated in later chapters, especially in Chapter 4, it is to be noted that using a humanoid robot pointing like a human can be quite helpful as an observer may be familiar with such type of pointing. Another type of robot will be a robotic arm which, similarly to humanoid robots, exhibit a physical presence and allow a limited set of gestures to be produced. However, the nature of the portrayed gestures may be quite different and therefore may be a contributing factor to the interaction.

Having decided on what types of robots will be most suited for the task it is important to identify the type of actions that these robots will be carrying out and also the demand on the tele-operator. Robots can interact at different levels in the context of the investigation. At a basic level of interaction, they can remain static while reproducing the voice of the remote operator. The usefulness of the robot in such an interaction is rather limited as, although it creates a physical presence with which the co-located users can interact with, it will not provide much information for the group interaction. On the opposite side of the spectrum, the robot could reproduce every action that the tele-operator is performing which poses some problems with respect to object selection and interaction. For the purpose of the investigation, only specific gestures will be reproduced, as a first step towards gesture reproduction the focus will be drawn to deictic (pointing) gestures. This type of gesture has already been proved successful when used in both real-world situations and in virtual environments. Expanding from this class of gestures a broader set of gestures which is with respect to Human Tabletop Interaction (HTI), a branch of Human Computer Interaction (HCI) primarily focussed on interaction between humans and TIDs, is going to be used. This limitation in gesture set is applied in order to allow only a fixed number of gestures to be reproduced and provide a specific scope for the investigation. Using gestures that would usually be used in HTI, allows for the focus to be drawn on implementing these gestures correctly and provide a foundation on which the rest of the study can be carried out.

Expanding from this class of gestures, a broader set comprised of a number of co-verbal gestures, is to be performed by a humanoid robot. While these gestures can be performed on their own, they are more useful when being performed in conjunction with speech. A humanoid robot performing co-verbal gestures has certain advantages as it allows processing of speech and gesture communication in a similar manner to face-to-face communication. However, in the case of a robot performing these gestures a number of issues can be raised. One of these issues is the synchronisation of the gestures with the accompanying speech, that is the gestures from the robot needs to be in line with speech. These will be further elaborated in Chapter 6.

In order to provide as little restriction as possible on the tele-operator, the tele-operator is going to be passively captured, that is, the capturing of the action of the tele-operator is carried out by the system itself, requiring as little as possible tele-operator active involvement in controlling the robot. This is more along the lines of the robot mimicking the tele-operator rather than having the tele-operator manually activating switches or operating a joystick to control the robot. As part of the overall system a gesture reproduction scheme, reproducing the captured actions directly, is favoured to a gesture classification scheme, that is, categorising the input gesture and then producing a pre-programmed gesture that would be appropriate to the interaction. This is in order to minimise the delay between capture and reproduction and also to allow a wider range of gestures to be used.

An important component of tele-operation that is beyond the scope of this work is the physical manipulation of objects. Firstly such interaction is not required for interaction with the Microsoft Surface (See section 3.3.1), and secondly, physical interaction with a real-world object would require adequate feedback to the tele-operator.

1.5 Contribution and Publication

During the course of this investigation the following papers were published:

- Hossen Mamode, H. Z., Bremner, P., Pipe, T., & Carse, B. (2012, December). Cooperative tabletop working for Humans and Humanoid Robots: Early investigations into artifact indication. In Cognitive Infocommunications (CogInfoCom), 2012 IEEE 3rd International Conference on (pp. 231-236). IEEE.

- Hossen Mamode, H. Z., Bremner, P., Pipe, A. G., & Carse, B. (2013, May). Cooperative Tabletop working for Humans and Humanoid Robots: Group Interaction with an Avatar. In ICRA 2013, 2013 IEEE International Conference on Robotics and Automation (pp. 231-236). IEEE.

While the above publications are focused on different aspects of the current investigation, they have also provided insight into a number of related problems.

The contributions to knowledge including the solutions to the previously mentioned related problems, which have been encountered during the investigation, are listed as follows:

- The efficacy of robots as tele-presence devices in group and dyadic interactions.

- The effect of different robot form factors on robot avatar efficacy

- A control module that allows humanoid robots to point to artefacts in a human-like manner.

- A control module that enables recognisable gestures to be produced by a humanoid robot in a time sensitive context.

- A relatively simple and easily moved capture system for the gestures of a participant.

- A novel method that allows interaction with augmented reality objects.

1.6 Thesis Structure

This thesis is composed of the following chapters, each covering a specific area. These chapters are:

Chapter 1 Introduction, the current chapter, introduces the context of the investigation and provides the motivation behind it.

Chapter 2 Background and Issues provides background information about the fields of research related to the current investigation with together with a critical analysis of the work carried out previously and leading to the enumeration of the hypotheses that are going to be tested as part of the investigation.

Chapter 3 Hardware lists the different hardware components that are used as part of the investigation together with a description of those hardware and the reasons behind why they are selected in contrast with other available options. Brief descriptions of their use and of additional non-hardware components that are related to these components are also provided.

Chapter 4 Robot Pointing, the first chapter that deals with the actual work carried out in the investigation focuses on pointing carried out by robots paying special attention to how different types of robot pointing are perceived by human observers. This chapter also describes a special control module that will be used in later parts of the investigation.

Chapter 5 Robot Mediated Cooperation, a chapter that focusses on a main topic of the investigation, that is, on the impact of having robots acting as embodiments for the actions being carried out by the remote members on different groups have on the overall group dynamics. This chapter also focuses on how the use of pointing from different types of robot embodiments impact the group productivity.

Chapter 6 Robot Gesturing, is focussed on how effective the actual tele-operation mechanism, used in the current investigation, is at reproducing recognisable gestures, not limited to deictic gestures, from a tele-operator who is being passively captured. It includes a description of the tele-operation mechanism together with a description of the positions of the different components. It also acts as a precursor to the following chapter which actively uses the tele-operation mechanism as part of a group interactive activity.

Chapter 7 Robot Presence, the final chapter that deals directly with the experiments that are carried out as part of the investigation, is focussed on the different types of presence that are exhibited by robotic embodiments in tele-presence settings. This is carried out with special attention drawn to the social presence aspect of the interaction and how different types of robot impact the type of presence exhibited.

Chapter 8 Conclusion, the final chapter of this thesis, summarises the findings of the investigation and provides a discussion on how these findings relate to the questions and the answers that are provided for the questions. It also provides a brief description of further work that could be carried out after the investigation.

1.7 Chapter Conclusion

The purpose of this chapter was to introduce the subject matter of this thesis. This was carried out by providing a brief historical setup of the two main elements of the investigation which is Tele-presence Robotics and Social Presence in Tele-presence environments. These topics will be further described in Chapter 2. Also the motivation behind the thesis together with accompanying aims and objectives of the thesis were stated. Moreover, the context and scope of the thesis were described with special attention as to where the focus of the research, carried out as part of the thesis, was aimed at. Finally, a brief description of the general structure of the thesis was presented detailing the purpose of each chapter.

Chapter 2 Background and Issues

2.1 Chapter Introduction

In this chapter, some key elements are introduced, while the main concepts are defined, previous works that are related to these concepts are also discussed. As outlined in Chapter 1, the focus of this thesis is on issues relating to Tele-Presence Robotics and Distributed Group Interaction. The related works therefore come from different established fields of research such as Human-Robot Interaction (HRI) and Computer Supported Cooperative Work (CSCW). Both of these fields, while being independent fields of research and addressing core issues that are specific to them, can also be viewed as being part of, related to or even, in the case of HRI, an evolution of Human-Computer Interaction (HCI), and such share certain common aspects and issues.

In this chapter brief descriptions of both HRI and CSCW are provided along with discussions on related works that is of particular interest in the context of this thesis. Since the type of CSCW that is being implemented in this thesis is using a Tabletop Interactive Device (TID) to support the Distributed Group Interaction, certain elements of Human-Tabletop Interaction (HTI) which is a form of HCI that is mainly with how humans used TIDs to support their work either locally or in a distributed setting will also be discussed. Finally, these discussions will also include gesturing, a type of interaction, that falls under the field of Human-Human Interaction (HHI), as it is going to be used as part of the interaction through the robot avatar in the Distributed Group Interaction.

Figure 2.1-A depicts a brief overview of the thesis with respect to the other fields of research mentioned above.

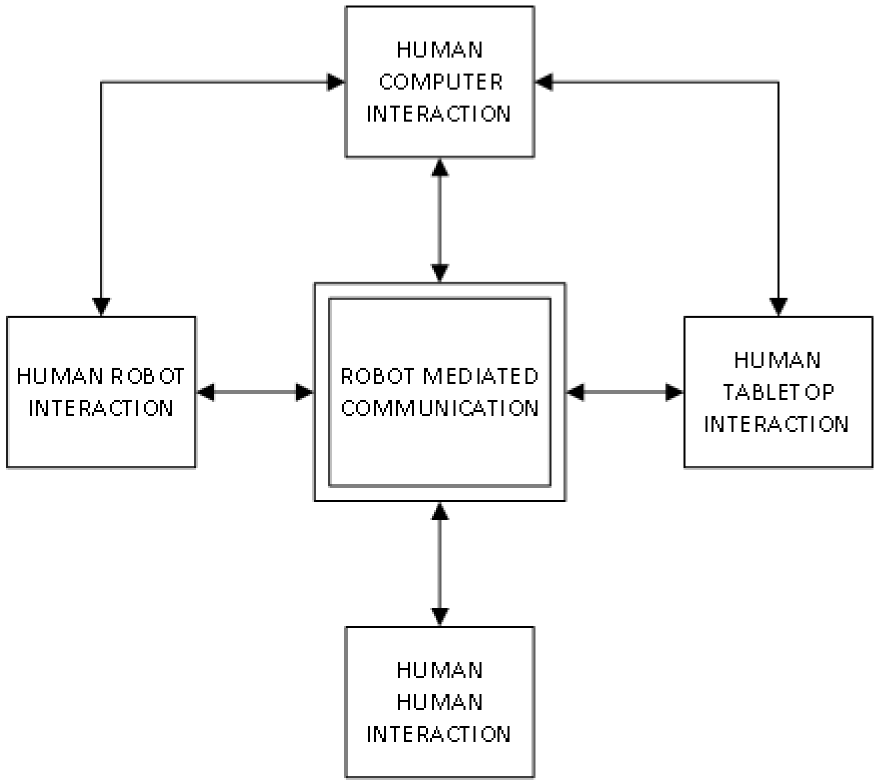

Figure 2.1-A: Thesis in Context

2.2 Interaction and Related Topics

While the two main types of interactions that are being studied in this thesis were briefly introduced in Chapter 1, it is in the following subsections, that the associated fields of research are described with respect to this thesis. In the following subsections, each of the different types of interactions between the elements and entities that make up the thesis are explained. Since Human-Robot Interaction (HRI) is the main focus of the thesis a brief description of it is first established which is then followed by sections describing two main topics that are fall under HRI but are of paramount importance to this thesis, Tele-presence Robotics and Humanoid Robotics. Since Human-Human Interaction (HHI) is quite a wide field of study the following two sections deal with specific elements of it that are used in this thesis, that is, Gesturing and specifically Pointing Gestures. Finally while this thesis is not directly going to delve into issues that relate to Human-Tabletop Interaction (HTI) the fact that TIDs are being used to facilitate group interactions HTI is going to be mentioned in the final topic in this section which deals with CSCW.

Figure 2.2-A depicts the types of interaction mentioned above together with the elements of the scenario that they affect and also the relationship between the different elements.

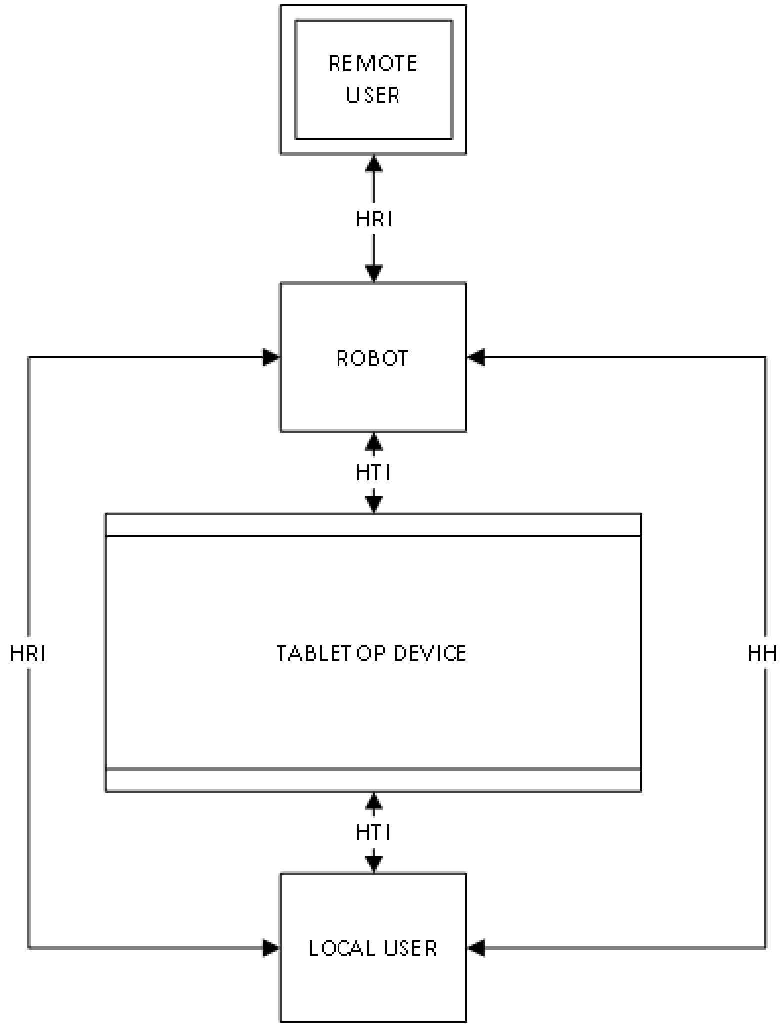

Figure 2.2-A: Interaction Types and Use

Table 2.2-I: Interaction Types

HHI	Human-Human Interaction
HRI	Human-Robot Interaction
HTI	Human-Tabletop Interaction

2.2.1 Human-Robot Interaction

In the definition put forward by Drury, Scholtz, and Yanco, for the term Human-Robot Interaction (HRI), it is defined as the study of humans, robots and how robots can be made accessible to humans (Drury, Scholtz, & Yanco, 2006). While HRI is related to Human-Computer Interaction (HCI), it also involves other fields that are not often present in traditional HCI such as robotics and artificial intelligence. Moreover, HRI has raised some issues that are not commonly associated with HCI, one of which as pointed out by Young, Hawkins, Sharlin, and Igarashi being the safety of the user (Young, Hawkins, Sharlin, & Igarashi, 2009).

According to Goodrich and Shultz, HRI is also markedly different from HCI in several ways namely (Goodrich & Schultz, 2007):

- the levels of human interaction involved,
- the interaction with the environment of the user,
- the possibility of hardware malfunctions during operation,
- the environment in which interactions occur.

With the increasing use of robots in settings other than factory floors, where they were programmed to perform fixed repetitive tasks, as agents that assist people in their day to day lives has caused a redefinition of the type of interaction carried out between humans and robots. This new type of interaction, as well as the social aspect of it, has been pointed out by Bartneck and Forlizzi among others (Bartneck & Forlizzi, 2004). While the new type of interaction between humans and robots can be viewed as either being tele-operation or direct interaction (Goodrich & Schultz, 2007); several taxonomies have been put forward to classify the new types of roles that humans can fulfil with respect to robots (Goodrich & Schultz, 2007). One of these taxonomies, developed by Yanco and Drury, defines the roles of humans in HRI as being (Yanco & Drury, 2004):

- Supervisor, where the user interacts with robots by issuing high level commands to them;
- Operator, where the user directly commands the robot to perform specific actions;
- Mechanic, where the user commands the robot while monitoring the status of the robot;
- Teammate or peer, where the user interacts with the robot as a member of a team;
- Bystander, where the user has a passive role during the interaction.

It is to be noted that the above taxonomy is limited to the active use of the robot or just passively viewing what the robot is doing. Despite this limitation, this taxonomy can be used to define some of the roles that the humans are portraying in the later part of this thesis.

2.2.2 Tele-presence Robotics

The term tele-presence robotics as presented by Goodrich and Schultz is defined as the use of robotic technology in order to allow users to have a tele-presence experience (Goodrich & Schultz, 2007). In the work of Kidd and Breazeal, tele-presence was defined as the use of technology in a way that allows users to be able to interact with an environment which is at a remote location with respect to where the users are physically present and with any object present in it (Kidd & Breazeal, 2004). This tele-presence involves the use of sensors in order to sense the environment, demonstrate a presence via an embodiment or either directly or indirectly affect the remote objects (Kidd & Breazeal, 2004). Though it is primarily used in situations where the environment is physically remote to where the user is; tele-presence robotics is also used where the environment in which the actions are to be performed are not suitable for the user to be physically present in (Bartneck & Forlizzi, 2004).

Tele-presence robotics has been used in a number of implementations; for example Ortmaier and Hirzinger, mentioned the use of tele-presence robotics for surgery in the form of tele-surgery, where a surgeon is able to operate on a patient from a remote location (Ortmaier & Hirzinger, 2000); another field mentioned by Goodrich and Schultz was space exploration, where the operator can command robots that are moving in an environment that is unsuitable for human presence (Goodrich & Schultz, 2007); or remote communication, where robots are used as a presence in a remote location. As pointed out by Goodrich and Schultz tele-presence robotics allows the synergy between human intellect and know-how, and robots; it is therefore critical in situations where the environment provides inadequate support for humans to be physically present (Goodrich & Schultz, 2007). This concept of providing a physical presence through the use of tele-presence robotics can be further expanded to include the use of such technology in areas where humans can be physically present, but are not able to do so in a timely manner, even into providing a physical aspect to video conferencing.

2.2.3 Humanoid Robotics

According to Goodrich and Schultz, the term humanoid robotics can be defined as the part of robotics that deals specifically with robots that possess anthropomorphic dimensions, possess human-like form or features; and behave in a human-like fashion, and in the case of autonomous robots also possess human-like reasoning (Goodrich & Schultz, 2007). The use of the humanoid form presents a number of advantages and has been used in a number of different works among which is the work of Billard, Robins, Nadel, and Dautenhahn, where small humanoid robots were used to teach basic gestures to autistic children (Billard, Robins, Nadel, & Dautenhahn, 2007). One of the elements of importance in that study was that humanoid robots were used as their behaviour can be identified as easily as human behaviour (Billard, Robins, Nadel, & Dautenhahn, 2007). This usefulness of the humanoid form for robots, with respect to interactions with humans, has been argued to be attributable to either a humanoid form or the presence of features that provides specific visual elements (Bargh, Chen, & Burrows, 1996). According to this line of reasoning, a person interacting with a humanoid robot may be able to use the same mental processes for understanding communication and motion that they usually apply to the interaction with another human being performing gestures to that of humanoid robots (Bargh, Chen, & Burrows, 1996). A different explanation, for the findings of Billard et al., can be found in the work of Oberman, Pineda, and Ramachandran, where comprehensive gestures from robots were found to be quite useful since they can trigger the mirror neuron system allowing for gestures to be more meaningful (Oberman, Pineda, & Ramachandran, 2007). Consequently, while the presence of features and the form of the robot may have an impact on the interaction it is their contribution to the gestures being produced that is of prime importance here. It is important to note that the term "humanoid robot" defines the form and features of the robot not the physical size of the robot, which may also have a potential impact on its use.

When referring to the use of the humanoid form of a robot for a task where it needs to interact with humans, one element to keep in mind is the acceptance of the robot by the human it is interacting with. One work that is often cited when dealing with the acceptance of a robot with respect to its form is the work of Mori (1970). In his work, it was noted that the form of different objects, some of which sharing human-likeness, affect how likely they are to be accepted by humans. While there is an initial trend whereby the more human-like the object is the more likely it is to be accepted, it is the sudden and acute dip in acceptance when an object looks too human that is of high importance. This dip is referred to as the 'uncanny valley' since the trend resumes when the object becomes similar to a human being. The work of Mori was further extended by MacDorman and Minato where this effect was seen to be further accentuated by the movement from the object, making the acceptance or rejection more pronounced when the same object is moving compared to when it is still (MacDorman & Minato, 2005; Mori, 1970). The findings of MacDorman and Minato is depicted in Figure 2.2-B.

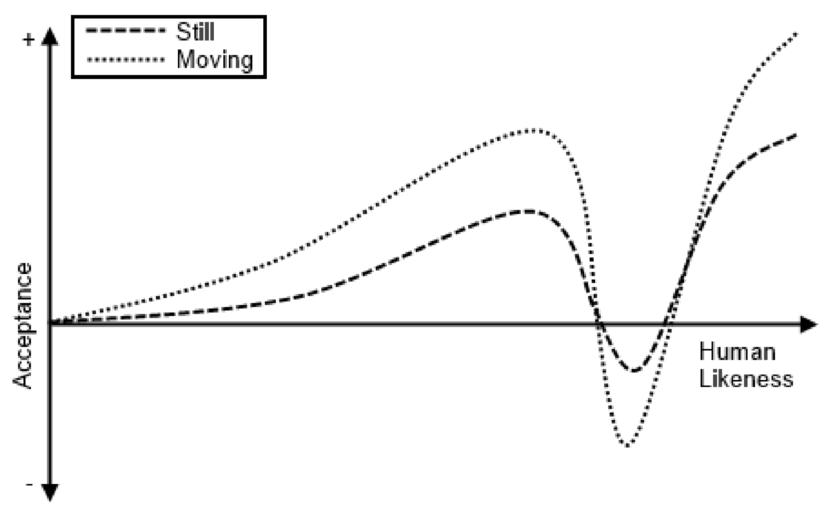

Figure 2.2-B: The Acceptance scale with the Uncanny Valley

The issue of physical form with respect to the acceptance and usefulness of robots for interaction is not limited to the still or moving forms as described by MacDorman and Minato (2005). According to Kose-Bagci, Ferrari, Dautenhahn, Syrdal, and Nehaniv, except from the physical form of the robot, the behaviour of the robot also plays an important part in the acceptance or rejection of the humanoid robots (Kose-Bagci, Ferrari, Dautenhahn, Syrdal, & Nehaniv, 2009). The idea of using social cues together with humanoid robots, according to Goodrich and Schultz among others, may lead to the implementation of social robots that may assist people with their tasks (Goodrich & Schultz, 2007).

Although not the focus of this thesis, one important use of humanoid robots is to be able to physically interact with objects designed for humans. In the work of Morales, Asfour, Azad, Knoop, and Dillmann, it was shown that a humanoid robot possessing humanoid hands can be used to manipulate objects in the real world (Morales, Asfour, Azad, Knoop, & Dillmann, 2006). One of the advantages of having a normal sized humanoid robot is that it is suitable to interact in environments that are usually designed for humans; apart from the work of Morales et al. which was focussed on grasping of objects and direct manipulation, other works such as the one carried out by Hirai, Hirose, Haikawa, and Takenaka, was focussed on how to allow humanoid robots to move in environments designed for humans which may include features such as steps as part of staircases (Hirai, Hirose, Haikawa, & Takenaka, 1998; Morales, Asfour, Azad, Knoop, & Dillmann, 2006). One of the reasons for not using physical object interaction in this thesis is that in order to allow proper tele-presence object interaction a feedback signal will be required to prevent any discrepancy in the interaction.

From the works presented above, humanoid robots have been shown to possess certain advantages over other types of robots. Some humanoid robots may be suitable to move and interact with people in an environment primarily designed for humans. Humanoid robots have already been used as an appropriate medium for the representation of human actions, since the presence of certain features make the actions easily identifiable by human beings. Consequently, for the direct portrayal of actions from a human being, it is suggested that a humanoid robot would be a more effective avatar than other robot forms.

2.2.4 Gesturing

In the definition provided by Kirk, Rodden, and Fraser, for the term gesturing, it was defined as being the use of movement from any part of the body in order to express thought or help emphasize speech (Kirk, Rodden, & Fraser, 2007). Following from this definition any movement made with the purpose to communicate can be considered a gesture (McNeill, 1985), this includes both movement co-occurring with speech such as illustrators and those that occur without speech such as emblems. This is in contrast with other visible movements that may occur during discourse but do not convey meaning such as self-touching and object manipulation (McNeill, 1992). The classification of gestures into illustrators and emblems, proposed by Bavelas, Chovil, Lawrie, and Wade (1992), is mainly focussed with the use of hand gestures in and outside of conversation. According to this classification of gestures, Illustrators, hand gestures that are used during conversation can be subdivided into Topic, focussed on the content of the speech, and Interactive, which is for the benefit of the social aspect of the conversation rather than the words being spoken. While Illustrators are linked with speech they tend to be spontaneous and unique to the person and the situation, furthermore they can reveal the inner thoughts of the person (McNeill, 1992). Therefore the illustrators are gestures that emphasize what is being said and therefore it is unusual for these gestures to be used to provide a different imagery to what is being said (McNeill, 1985).

While the classification of Bavelas et al. (1992) is suitable for classifying hand gestures, other classification systems were also proposed by Efron (1941), Ekman and Friesen (1981), Freedman and Hoffman (1967), and McNeill (1992). According to McNeill's classification, which is going to be used in this thesis, gestures can be classified as one of the following categories:

- Iconics: these gestures are linked to the semantic content of the language

- Metaphorics: these gestures are abstract representation of the content

- Beats: these repetitive gestures are unlinked to the content of the language but are still used during a discourse in order to bring emphasis to a specific part

- Cohesives: these gestures are used to bring together different part of a discourse that are related

- Deictics: these gestures are used to point to either objects in the vicinity or completely abstract objects and will be further discussed in the following section

While the above classification is appropriate when used in speech, McNeill also identified three other categories of gestures that are not part of speech

- Pantomimes- Use of hand and face expressions to communicate an idea

- Emblems- Use of hand gestures that serve a specific function

- Sign language- Use of hand gestures in the absence of any speech as a form of communication

The ordering of the these categories has been defined by McNeill as part of his Kendon's Continuum and goes from a strong cohesion with speech to being completely independent of it (McNeill, 1992). This classification of gestures was selected as it encompasses most of the other categorisation schemes and also since it uses non-speech specific gestures, which are going to be referred to later in this thesis. Iconics and metaphorics in this classification method are closely linked with the illustrator gestures previously mentioned and are usually tri-phasic movements, the preparation, stroke and retraction while the beats are two phasic. The stroke part of the gesture being the part of the gesture that needs to be in temporal synchrony with the word being emphasized in the speech (McNeill, 1985). The synchronicity of gesture with speech can follow two rules:

- Phonological synchrony rule proposed by Kendon (1980):

According to this rule, the stroke part of the gesture precedes or ends before but not follow phonological peak of speech.

- Semantic synchrony rule

According to this rule if both speech and gesture are to co-occur they must cover the same idea unit.

As mentioned previously gesturing is a natural occurrence during speech, as it helps to support the same idea that is being conveyed using words (Freedman, 1977; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001), however it is not only the form of the gesture but also the frequency that can vary depending on the individuals and the situation (McNeill, 1992). For example illustrators have been shown to be affected by the perception of the accessibility of the gestures by the interlocutor, since in the works of Cohen and Harrison (1973) and Cohen (1977), illustrators were shown to be less used when the receiver was not able to see them. While using the above scheme, it is possible to categorise gestures with respect to their function, it is also to be noted that all gestures in a category can vary greatly in term of the size of the perceived motion. For example, McNeill (1985) pointed out that no two interactive gestures are exactly alike and according to Goodwin pointing can be accomplished in different ways, one of which being the use of gaze, involving a simple eye movement, or carried out using arms requiring the use of arms and hands and also a greater movement (Goodwin, 2003).

Gesturing is an important component of Human-Human Interaction and there has been several studies carried out by Cerf, Harel, Einhauser, and Koch, Emery, Imai, Kanda, Ono, Ishiguro, and Mase, Ito, Hayakawa, Hotokata, and Terada, Kleinke, Matsumoto and Zelinsky, Tan, Robinson, Culbertson, and Apostolopoulos., Wu and Huang among others in order to find out how to correctly read such gestures from humans (Cerf, Harel, Einhauser, & Koch, 2008; Emery, 2000; Imai, Kanda, Ono, Ishiguro, & Mase, 2002; Ito, Hayakawa, Hotokata, & Terada, 2003; Kleinke, 1986; Matsumoto & Zelinsky, 2000; Tan, Robinson, Culbertson, & Apostolopoulos, 2010; Wu & Huang, 1999). Other studies for example, the one carried out by Cheng and Kuniyoshi, were more concerned with how to successfully implement them in other embodiments (Cheng & Kuniyoshi, 2000). This is particularly important as the use of the wrong gesture may be misleading to a person and can therefore have detrimental consequences on the effectiveness of the overall group interaction (Kirk, Rodden, & Fraser, 2007). As pointed out by Goodwin, while the presence of particular limbs are crucial for some gestures to be correctly performed in the pre-established context, some gestures may require other elements present in the environment for them to be useful (Goodwin, 2003). Moreover, Breazeal pointed out that para-linguistic communication signals, gestures not meant to completely replace the use of verbal communication, as opposed to gestures used in sign language, can be used to enhance the audio information (Breazeal, 2003).

Since they form a crucial part of interaction between human beings, gestures have been used to help communication in other media than the real world. Using virtual embodiments in virtual environments, gestures have been proved to facilitate communication, with varying success (Heath & Luff, 1991; Kirk, Rodden, & Fraser, 2007).

Following from the above ideas, it is suggested that humanoid robots in particular, or other robots that possess the required elements to fulfil the requirements of the gestures, would be able to make use of gestures to enhance the interaction with human beings. It is suggested that in the context of social presence, such as, the focus of this thesis, these gesture-enhanced robots would even be able to adequately convey the presence of a remote participant to the participants that are in the immediate vicinity of the robots.

2.2.5 Pointing Gestures

In the previous section, one class of gestures that was identified is the deictic gestures. These deictic gestures are often commonly known as pointing gestures. Pointing gestures is an important group of gestures which is markedly different from both the super-category of illustrators, including Iconic and Metaphoric gestures, which while helping with speech are not critical for the comprehension and emblems who have no speech counterpart (Ekman & Friesen, 1981). As mentioned by Levelt, Richardson, and La Heij (1985), deictic gestures are almost critical in the proper comprehension of the part of speech that use the accompanying deictic terms, since a far less efficient mode requiring more words to convey the same meaning can still be used (Bangerter, 2004).

Although pointing gestures are critical accompaniments to some specific deictic terms there are myriad ways in which the pointing can be performed (Goodwin, 2003). These pointing gestures vary both in terms of form and the visibility of the gesture (Goodwin, 2003). However, what does not vary is the purpose and basic elements of the pointing. As defined by Kita, pointing is a communicative body movement that uses a vector projected from a body part to designate a general direction, a specific location or a designated object. Since its development in early infancy (Masataka, 2003) as part of word-referent (designated object), the pointing gesture quickly becomes part of the day-to-day interaction with others (McNeill, 1992). Some researchers, notably Butterworth, have argued that pointing as part of a communicative process is uniquely a human trait as pointing in this form is not present in other primates (Butterworth, 2003). Since it is multi-purpose in the object it is referring to, pointing can also be used to create iconic representations by defining limits or points of the other gesture (Kita, 2003).

One important element of pointing is the form of the gesture. While the most common version of pointing, that is using the index finger, is usually easily recognised in most parts of the world (Kita, 2003), it isn't however universal as argued by Wilkins (2003). According to Wilkins, while there are cultures that don't use index for pointing to refer to objects, nevertheless there is a culture specific gesture that replaces the index pointing gesture (Wilkins, 2003). On the opposite end of the spectrum, there are animals such as apes and dogs (Hewes, 1981) that tend to produce a pointing like gesture but these are not innate behaviours but are instilled in the creatures through training and are not part of a discourse and therefore lack the proper context to be classified as a gesture.

Finally, as mentioned previously, pointing is not only performed using the index finger, other body parts of the body can also be used for pointing. These other pointing gestures work on the same principle of creating a vector to direct attention to the object being designated (Kita, 2003). In the case, of the speaker holding an object in the pointing hand it is not uncommon for the person to use the object (Goodwin, 2003). Moreover, other more subtle ways of pointing can still be used to cause a shift of attention, such as the use of gaze (Butterworth & Itakura, 2000). Using gaze with regards to pointing can act in two different ways, firstly as a pointing gesture in its own right, where gaze is used alone to direct attention in a subtle way or as supportive of more visually indicative pointing gestures (Kita, 2003). Using gaze on its own for pointing is less used since it is hard for the interlocutor to create a vector when the lever is so small, limiting the gaze to designate wide areas of attention (Butterworth, 2003). However, when used in tandem with finger pointing, the gaze can actually help direct attention from the interlocutor to the designated object more easily (Kita, 2003).

2.2.6 Computer Supported Cooperative Work

Up until now, the concepts described in this chapter have been focussed on robotics and the types of interactions that are being studied in this thesis, however in this section two different concepts that, while not directly related to interaction, are still of importance for the thesis and are going to be described. The first of these concepts is Computer Supported Cooperative Work (CSCW) which is defined by Brown and Bell as the use of computers and associated technologies to help groups of people to work more efficiently together (Brown & Bell, 2004). A similar definition for CSCW was also used by Rama and Bishop (2006).

The second concept in this section is groupware and is usually used in conjunction with CSCW; although CSCW can be established without groupware, groupware is usually used to facilitate CSCW. This close connection between the two concepts has led to those concepts being used interchangeably. However according to the definition provided by Brown and Bell, groupware is defined as any software specifically designed for use by groups of people working together (Brown & Bell, 2004).

Relevant CSCW applications are further elaborated in later chapters, here two examples are given for clarity; firstly, when the members of a group use the TID to collaborate with each other and secondly when the members of the group interact via the robotic embodiments. It is to be noted that CSCW is not limited by the members of a group being co-located, that is, it is also applicable in cases where groups are comprised of remote members (Tuddenham & Robinson, 2007; Tuddenham & Robinson, 2009).

2.2.7 Section Conclusion

In this section, the focus has been on providing a broad description of the related areas of work. Part of this description included the clarification of certain concepts that are of particular interest in the current investigation. These concepts included both the different interaction types that have been identified as forming part of the interaction between geographically dispersed groups and also technologies that may be used to support such interactions. While a broad view of the concepts was provided in this section together with the ways in which they are related to this thesis, these concepts will be further discussed in the following section. Together with the discussion of the different concepts, a review of the related works in each discussed area will be provided to create a proper context for the work that is presented here.

From the description provided for the different topics that have been identified as being part of the related topics, a few important findings are:

- Tele-presence robotics can provide a physical embodiment for a remote interlocutor and can enhance traditional video conferencing.

- Gesturing is dependent on the presence of certain features and when taking the human mirror neuron system into consideration the proper robot moving in an appropriate manner may have a net gain in the perception of the gesture and consequently the overall group experience.

- While CSCW helps co-located individuals to work more efficiently together, this can also be applied, to a certain degree, to situations where the interacting individuals are geographically dispersed.

2.3 Related Previous Work

In order to provide a proper context for the work presented here, it was important to first understand and critically assess the work that has already been carried out in the related fields mentioned in the previous section. In this section of the chapter, the works that were analysed and found to be of particular relevance to the topic of the thesis are discussed. It is to be noted that some of the technology that is being mentioned in this thesis is relatively new, or has recently become accessible or affordable, and this is a reason for the scarcity of material that deals directly with the objectives of this thesis.

2.3.1 Distributed Group Interaction

Though Tabletop Interactive Devices (TIDs) have been around for a relatively short amount of time, they have been the focus of much attention. As mentioned by Tang, Neustaedter, and Greenberg, these devices, in their designed purpose, are meant to allow a group of people in an area to have a platform on which they can build and share ideas with other members of the group (Tang, Neustaedter, & Greenberg, 2006). Studies have been carried out, for example by Tuddenham and Robinson, to discover the effects of the use of such devices on the interaction between the members of a group (Tuddenham & Robinson, 2007). In the study it was found that TIDs helped co-located users to interact in a more meaningful manner allowing them to better utilise the available resources and increasing cooperation. They also investigated whether the benefits to a group, where the members are co-located, can also be extended to situations where one or more members are present in another location (Tuddenham & Robinson, 2007). They found that the same level of interaction, which was achieved when all the members were co-located around the TID, was not possible when the entire group was not co-located, despite each member still having access to TIDs. Moreover, it was suggested that the reason for this was a lack of presence of the remote participants in the locations that they were not physically present in.

Apart from investigating what the actual benefits to the distributed groups are when interacting through TIDs, studies such as the ones carried out by Bos , Shami, Olson, Cheshin, and Nan, Bradner and Mark, or MacDorman and Ishiguro were focussed on whether the presence of a person can be extended from one area to another using different types of technologies (Bos, Shami, Olson, Cheshin, & Nan, 2004; Bradner & Mark, 2002; MacDorman & Ishiguro, 2006). They found that one effect of the limited presence, generated by the remote participants in the group, was a split in the group interaction. That is, the co-located members of the group segregating the remote participants out of their group; hence, the remote participants react by working exclusively with other remote participants. This emphasizes the need for an adequate level of presence to be generated for the remote members in the co-located area.

The concept of allowing members of a group to have a presence in another area to where they are physically located is not solely linked to TIDs. Studies carried out by MacDorman and Ishiguro, and Tang et al. were focussed on finding out how different types of embodiments can transfer social cues, important elements of social interaction (MacDorman & Ishiguro, 2006; Tang, Neustaedter, & Greenberg, 2006). From these studies, it was found that while physical representations of remote participants' actions are very useful, the use of virtual representations, in the absence of availability of a physical representation, can still have some limited usefulness. While having one or more robots display social presence on behalf of remotely located members of a group may be new, as pointed out by Bartneck and Forlizzi, the idea of having a social presence and the use of social cues in a different area than where the participants are physically present is not (Bartneck & Forlizzi, 2004).

On the subject of social cues and presence from embodiments, studies on physical embodiments have been rather limited possibly due to the high demand in resources of such endeavours. However, many more studies have been carried out using virtual embodiments that usually require less resources. In a study carried out by Hindmarsh, Fraser, Heath, Benford, and Greenhalgh, the same distributed group interaction problems, previously encountered, were extended to a different environment so that the interactions were with objects present in a virtual environment instead of those in physical environments (Hindmarsh, Fraser, Heath, Benford, & Greenhalgh, 1998). It was found that while virtual environments created some problems with respect to limited perception of the environment, group interaction was still achievable in a limited way.

Even in computer games and such virtual environments, as pointed out by Brown and Bell, the proper representation of social cues plays an important role in the use of avatars, virtual representation of people, or even non-playable characters (Brown & Bell, 2004). The proper representation of social cues, such as pointing, has been the focus of many studies for example the one carried out by Hindmarsh et al., which involved both virtual environments and the physical world (Hindmarsh, Fraser, Heath, Benford, & Greenhalgh, 1998). Furthermore, a study carried out by Drury, Scholtz, and Yanco (2003), has shown that having an embodiment, even a virtual one, displaying social presence to other members of a group can have a positive impact on the group interaction.

From the studies described in this section, it was found that the nature of an embodiment, physical or virtual, the presence generated and the related behaviours, including social cues, are important factors that affect distributed group interaction. The implications for this thesis is that in order to investigate the use of robots as avatars for remote users, it is important to first test how different robotic embodiments affect group interaction. Secondly, to investigate what kind of presence is generated by the different robotic embodiments, and finally the effect of social cues from robots on the group interaction.

2.3.2 Robotic Tele-presence

The concept of tele-presence, as defined by Kidd and Breazeal, is concerned with establishing a presence at a different location to where a user is physically present; it involves either sensing the remote environment, providing an embodiment for others to interact with or performing actions in the remote environment (Kidd & Breazeal, 2004). For the purpose of this thesis, it is the last two parts of the definition that are going to be of great importance. In this thesis, the embodiments that were implemented made use of robots that generate a physical presence in the real world. However, other studies have been carried out, such as the one carried by Brown and Bell, where the nature of the embodiments was virtual and these were used to study social interaction and cues in virtual environments, specifically game settings (Brown & Bell, 2004).

The concept of a remote embodiment, with respect to a remote user, is to provide a presence in the environment of a local user with which the user can interact. Having a robotic embodiment can affect the local environment in a number of ways; the mere presence generated by a robot acting as an embodiment may cause an impact on the interaction and make the local users interact with the remote person in a different way than if there was no physical embodiment (Guerin, 1986). One suggested explanation for this behaviour, in the case of a single embodiment, is that the embodiment used, either being fixed or mobile, restricts the perception of the remote person in the local environment to one position. With the remote individual effectively localised to a specific position, in the local environment, the local members are then able to direct their attention to where the remote user is perceived to be situated (Sakamoto, Kanda, Ono, Ishiguro, & Hagita, 2007). This concept of a local position with respect to remote participants was again pointed out by Tang et al., in their study with projected virtual arms representing the arms of the remote participants, where it was used as a critical element of group interaction (Tang, Neustaedter, & Greenberg, 2006). In the work of Hindmarsh, Fraser, Heath, Benford, and Greenhalgh, where the focus was on interacting with objects, albeit virtual ones, it was found that it is important for the participants to know exactly where the object is situated relative to the participants (Hindmarsh, Fraser, Heath, Benford, & Greenhalgh, 2000). In that study, it was suggested that having this type of knowledge can allow for actions to be directed to a particular object and can therefore help the collaboration become more efficient (Hindmarsh, Fraser, Heath, Benford, & Greenhalgh, 2000). This perceived presence of objects, even in virtual environment, is not limited to only inanimate objects and can also be applied to cases where a person is interacting with another human being. These localisation effects are of particular importance to the work presented here, as it is suggested in this thesis that, making use of these effects, it is possible to enhance the interaction with remote participants. For example, having a robot localising a remote individual in the area of the robot will allow the local individuals to interact with a local presence. This local presence, even without the use of movement and gestures, will provide benefits as the local individuals will be able to address and query the remote person directly as he would be effectively occupying a fixed point, that is, embodied by the robot in the environment.

According to Kidd and Breazeal, the presence of an embodiment in the local environment of a user apart from providing a presence, with which the local users can relate to and to which conversation or gestures can be directed to, also affects the way in which actions are perceived by the local users (Kidd & Breazeal, 2004). One important element mentioned by Goodwin with respect to the perception of a hand pointing gesture is that the trajectory of the motion, while the action is still being carried out, in conjunction with the pose of the body tend to hint to what action is being carried out or where the focus is going to be directed (Goodwin, 2003). Tang et al. pointed out that this trajectory based motion is an important element of efficient HHI and is based on knowing where the remote user is meant to be locally present with respect to the local users (Tang, Neustaedter, & Greenberg, 2006). In the work of Tang et al. that information was limited to a planar view, restricting that perception to a two dimensional setting, however the use of a robotic embodiment would allow this important trajectory information to be reproduced in a three dimensional setting, thus conveying more information (Tang, Neustaedter, & Greenberg, 2006).

While pointing is a useful gesture with respect to remote interaction via an avatar, it is not the only gesture that can be useful during avatar mediated interaction. During the human to human interaction processes, social elements such as cues and gestures are used to support or give emphasis to an argument or to regulate the interaction (Kendon, 2004). These gestures that do not form part of the pointing class of gestures are nevertheless quite important for use on an avatar if a proper presence is to be exhibited by the avatar.

However, the inclusion of other types of gestures raises certain issues especially when these gestures are to be performed by a real world avatar such as a humanoid robot. In the work of Hindmarsh et al., it was stated that whether gesturing is being performed in a real or a virtual environment, special care needs to be taken to provide not only the right gesture but also to do so in a manner that the interlocutor can understand (Hindmarsh, Fraser, Heath, Benford, & Greenhalgh, 2000). It may therefore not be possible to directly apply these captured gestures from the remote operator of the robot. In order to maintain the context and allow correct perception of the gestures from a remote operator of the avatar, the captured gestures may need to be transformed, for them to be correctly interpreted by the local interlocutors. In other words the form of the gesture may need to be altered while preserving the essence of the gesture.

Optimally, it would be quite useful to be able to use direct mimicking, that is, whatever the remote person is doing is just copied to the local environment as was conducted by Cheng and Kuniyoshi (2000). As pointed out before the use of such an approach with robotic avatars raise a few issues, one of which being the physicality of the robot. Using a robotic avatar, especially a full size humanoid, may first of all raise issues with regards to safety. If a robot is to just copy the actions of a remote person any rash action by the remote operator will also be transferred and may cause damage to objects in the local environment. However, finding a robot that can move in the exact same way and at the same speed as a human may not be possible. Using a robot with slower speed can raise issues such as delays in action execution due to the delay of moving from rest or the robot not reaching the target position before the following command is sent. While this may make the robot unstable, as in the exact position of the robot at any particular point in time cannot be guaranteed, a related problem, may make it unusable. This related problem is the loss of temporal synchronicity between the gesture and speech as described by Cassell (2000).

Another approach than just mimicking the action directly, mentioned in the work of Riek et al. which involved life size humanoid robots, is having the actions carried out by the embodiment in a manner that is consistent with the actions while still taking into account the capabilities and limitations of the embodiment (Riek, et al., 2010). Using this approach, where the embodiment is to carry out the actions in a way that is specific to itself, the way in which the actions are to be portrayed is important. It has been shown in the work of Riek et al., that gestures do not have to be exactly human-like to be useful to a human observer (Riek, et al., 2010). It can therefore be deduced that it is more important for the actions to be appropriate with respect to the task at hand, than it is for them to be exactly human-like.

One possible way in which this can be achieved is through the use of gesture classification, that is, the gesture from a remote operator is first analysed and a suitable gesture that the robot is able to achieve is done in its place. This approach raises a number of issues when it is being performed as part of a live interaction, one of which being the time delay of the classification process and another being the over quantisation of the gestures. Another possible option, that is going to be discussed later in this thesis, is the use of a gesture modification scheme, which attempts a similar process but without the use of a classifier system. This gesture modification scheme attempts to modify the gesture while preserving the context of the gesture. One advantage of this option is the quantisation issue will not be present, that is, a gesture will not be required to already be known to the system for it to be captured. The amount of gestures being reproduced using the second method is likely to be much higher as unrecognised gestures will still be performed by the robot.

The focus of this section has been on the use of robotic embodiments in the context of tele-presence. After providing a basic definition of the concept of robotic tele-presence and how it relates to the focus of this thesis, the advantages of the use of such embodiments were described. The discussion then led into possible disadvantages or potential problems that may be encountered if robotic are used as embodiments for remote operators. As part of this discussion, potential solutions were identified and briefly described. These potential problems and their accompanied solutions, in particular how to allow the useful reproduction of gestures with a robot that may not be able to achieve similar movement speed as the operator will be the focus of later discussions.

2.3.3 Remote Gesturing

In the definition given earlier in the chapter for tele-presence by Kidd and Breazeal, it was mentioned that one of the three ways in which tele-presence can occur is when an embodiment is used to convey the actions of a remote user who is present in a different environment (Kidd & Breazeal, 2004). These embodiments can interact with objects in two ways, they can either interact directly with objects through the use of touch as mentioned by Bartneck and Forlizzi (2004) or refer to objects without actually making contact with them using different types of gestures as carried out by Kuzuoka, Oyama, Yamazaki, Suzuki, and Mitsuishi (2000). In this thesis, the focus is on the latter, that is, information conveying through gestures.

While these types of interactions are different, they are not mutually exclusive, in fact, gesture based interaction can enhance an experience that already involves direct object interaction by a robot. For example, in a setup where a robot is acting as an avatar for a remote user in a tele-presence setting and manipulating objects, it may also be useful to have gesture interaction, in order to increase the social presence of the remote participant. According to a study carried out by Liu, Laffey, and Cox, it was found that being part of a group assigned to a particular task usually entails not only working to complete the task but also interacting with the other members of the group in order to coordinate the effort and so be more efficient (Liu, Laffey, & Cox, 2008). Collaboration and coordination usually require communication between the different members of the group and while audio communication can help in coordination, Kirk et al. found that it is not as effective as the inclusion of gestures as part of the interaction (Kirk, Rodden, & Fraser, 2007).

As mentioned previously, due to the wide definition of gestures, there are different types of them. Since the use of gestures is context specific, meaning that the action being performed is linked to the situation that the gesture is taking place, to be used properly the gesture being produced by an avatar needs to fit the current task (Bartneck & Forlizzi, 2004). While some gestures can be easily performed by humans when these same gestures are being performed by an avatar in a different environment, as pointed out by Hindmarsh et al., poses certain new challenges (Hindmarsh, Fraser, Heath, Benford, & Greenhalgh, 1998). Even relatively simple gestures such as using a finger to point towards an object to draw focus can become quite complex when it is to be reproduced remotely as was shown by Kirk et al. (2007).

While a simple act of pointing is usually performed between at least two entities, one of which is using the gesture to draw the attention of the other to a specific area for cognition and action as described by Goodwin (2003), having to point to an object that is outside the visual range of the one doing the pointing can become quite complex. This becomes even more arduous when all the different requirements for proper pointing are taken into account:

- a body visibly performing an act of pointing;
- talk which both elaborates and is elaborated by the act of pointing;
- the properties of the space that is the target of the point;
- the orientation of relevant participants toward both each other and the space that is the locus of the point;
- the larger activity within which the act of pointing is embedded;

The use of gestures, in another environment that the user is present in, is not limited to the field of robotic tele-presence, gestures have been used in virtual environments for some time, such as in the work of Brown and Bell where the virtual environment in question was a game (Brown & Bell, 2004). The use of simple pointing in such settings has proved to be quite useful for coordination between the different members of a group in helping to avoid confusion that may arise from the use of audio alone as was pointed out by Hindmarsh et al. (1998). While these pointing gestures have similar requirements, as mentioned above, the benefits that they provide have been crucial. It is to be noted that in these examples, one of which being Brown and Bell, the inputs, that are being provided by the users as commands for the avatars to perform specific actions in virtual environments, are not the same gestures that are being carried out in the virtual environment (Brown & Bell, 2004). This leads to a situation of Representation-bridging communication as described by Baranyi and Csapo (2012).

While the correct gesture in a particular context can be very useful for coordination, the opposite has also been proven by Emery among others (Emery, 2000); the results of wrongly performed gestures or wrongly interpreted gestures can lead to confusion and frustration so it is important that the gesture is performed in an appropriate manner and for it to be understood properly by the interlocutors (Emery, 2000).

One important element when dealing with robot gesturing is the degree of autonomy on the gestures that are being carried out by the robot, that is, whether they are coming from the robot itself or from a remote operator. Firstly, in the work of Takayama, Groom, and Nass (2009) a toy robot is used to point at different objects in a social presence experiment, but the robot, used in that experiment, was producing the pointing from the commands received instead of the operator manually guiding it. However, in the work of Kuzuoka et al., where a robot is used as a mentor, the pointing gestures provided by the robot acting as an avatar were being actively coordinated by the remote operator through the use of a joystick (Kuzuoka, Oyama, Yamazaki, Suzuki, & Mitsuishi, 2000). This active control has also been carried out with the use of custom built rigs such as the one used in the work of Adalgeirsson and Breazeal (2010) where the gestures being performed by the robot avatar are being actively captured from the remote operator. Finally, a different form of control again for use on avatars has been used extensively in the animation and entertainment industry. This form of control does not make use of any handheld devices but instead uses the body of the remote operator as the source of the motion for the avatars.

As mentioned previously, these issues regarding the execution of remote gestures are particularly important in the case where a humanoid robot is acting as an avatar as is the proposed case in this thesis. As introduced earlier, the concept of the uncanny valley, when applied to humanoid robots require them to behave in a specific way, or risk a negative acceptance effect. This behaviour needs to be perceived to be acceptable for humans, or they will not be able to exploit the benefits that usually come with tele-presence and even cause some confusion. This requires the robots, in such a setup, to display social cues in a manner similar to how a local person would do it, for example the coordination of both gaze and finger movement when pointing to an object. While the previous works have mentioned that social cues are an intrinsic part of human-human interaction even across different media, the idea of adding coordinated cues to humanoid robots to emphasise the action being carried out is novel to this thesis.

2.4 Chapter Conclusion

2.4.1 Conclusion

The purpose of this chapter was to describe key concepts related to the thesis along with the presentation of some related works in order to put this thesis into context.

Analysis of the related work has led to the identification of key areas of research that can be built upon, as well as areas that have yet to be addressed:

- The way the nature of an embodiment impacts the perception of the action by a human.
- The impact of an embodiment representing the action of a remote user on group collaboration across vast distances.
- The impact of coordinated social cues from an embodiment on local users.
- The use of mimicry as a source of gestures for instruction coming from robot instructors.

2.4.2 Hypotheses

While the key concepts that are related to the thesis were described and discussed in this chapter, a number of issues were identified. In order to clarify those issues a number of hypotheses are going to be formulated.

One of the first issues identified from the literature, discussed earlier in the chapter, was with respect to distributed group interaction where it was found that the lack of a physical embodiment for a remote member of a group severely limits group collaboration. One important element of group interaction, either co-located or distributed, is group awareness, knowledge of what other people are doing (Gutwin & Greenberg, 1995); the lack of a physical embodiment limits that awareness both with respect to the position of the remote user in the local environment and the abilities of the user (Tang, Neustaedter, & Greenberg, 2006). While the lack of a physical embodiment does not prevent group collaboration, the collaboration within a group where all the members are not physically present tends to be severely affected (Bos, Shami, Olson, Cheshin, & Nan, 2004). The lack of a physical presence can create a presence disparity between the co-located members and the remote members leading to the remote members being either ignored or excluded from the group collaboration (Bos, Shami, Olson, Cheshin, & Nan, 2004). One possible solution to attenuate that problem is to create a physical embodiment for the remote users, since embodiments have been shown to have an effect on social interaction (Kose-Bagci, Ferrari, Dautenhahn, Syrdal, & Nehaniv, 2009). This proposed solution leads to the following hypothesis:

H 2.1 A physical embodiment of a remote user will allow for a higher level of interaction, that is, better collaboration and higher level of presence than when no physical embodiment is used.

While a physical embodiment is useful as a first step in establishing a presence for a remote member of a group; as mentioned earlier, the behaviour of the embodiment also plays an important part in the presence that embodiment exhibits. In previous work carried out in the field of remote presence, the use of gestures from robots have proved the usefulness of robot enabled gestures in group interaction and in exhibiting a presence on behalf of a remote participant (Adalgeirsson & Breazeal, 2010). According to Kose-Bagci et al., possessing an appropriate embodiment and supporting specific array of gestures play an important part in social interaction between a person and a robot (Kose-Bagci, Ferrari, Dautenhahn, Syrdal, & Nehaniv, 2009). However, while the use of robot avatars have helped in establishing a presence, the level of presence has been dependent on the level of supported gestures and have not been markedly different from co-located presence (Adalgeirsson & Breazeal, 2010). This leads to the following hypothesis:

H 2.2 A robot using an appropriate control scheme for portraying the actions of a remote participant in a group mediated task, will be able to provide a level of presence on behalf of the remotely situated participant to allow the group to behave and perform as if the group was co-located.

These main hypotheses are therefore proposed and will be tested in later chapters.

2.4.3 Summary

In order to investigate these hypotheses, a tele-operation system for a robot avatar is going to be implemented. The following chapter will therefore describe the components, both hardware and software, of this system. Following from which are chapters describing the experimental investigations conducted to evaluate its efficacy, and test the hypotheses.

Chapter 3 Hardware

3.1 Chapter Introduction

In the previous chapter, the context and underlying principles of this thesis, along with hypotheses to be tested were established. In order to investigate these ideas, a number of hardware components were required. In this chapter, the hardware components together with their associated software components are described. The description of the hardware components is going to include the capabilities of each hardware component together with its limitations. Also included in the descriptions are the reasons why it was selected for the system implemented in this thesis. Where appropriate, alternative hardware systems to those selected will also be discussed.

3.2 Robot Platforms

One of the most important hardware components of the system, which will enable the investigation to be carried out, is the robot platforms. As part of the investigation two types of robots are going to be used, a humanoid and an industrial robot arm.

In the World Robotics 2002 Report compiled by the U.N. and the I.F.R.R., three types of robots were identified (United Nations Economic Commission for Europe;The International Federation of Robotics, 2002):

- Industrial robots
- Professional service robots
- Personal service robots

The above classification is similar to the one put forward by Bartneck and Forlizzi (2004). According to which, industrial robots are those that are assigned a fixed repetitive task with minimal human interaction, while professional service robots are able to operate remotely in environments that may be potentially hostile to the human body. Finally, personal service robots are characterised by the direct interaction with human beings requiring them to behave in a socially acceptable manner (Bartneck & Forlizzi, 2004). Given the aim of this thesis, personal robots are proposed as most suited to the role of an avatar.

This classification is with respect to the services that the robot is to carry out, and hence does not restrict the form that the robot has to take. Although as will be shown later in this chapter, a large portion of socially interactive personal robots are humanoid. Conversely, industrial robots are usually task dedicated and generally provide little in terms of social interaction. This task focussed approach in industrial robots design allows them to be highly accurate and with high repeatability, however when they are to be used in social interactions they usually need to be supported by other means (Adalgeirsson & Breazeal, 2010). These two design paradigms are quite distinct from each other with the humanoid robots and their unique features geared towards social interaction while the non-humanoid geared towards task based efficiency. The division motivates the comparison between a humanoid and a more industrial design: do the social affordances of a humanoid make if more suited as an avatar than a pragmatically designed industrial arm. It is with this in mind that this section is going to be divided into humanoid and non-humanoid platforms.

3.2.1 Humanoid Robots

As mentioned previously, one important element for portraying gestures most notably the deictic gestures is the presence of certain specific body parts. While Hewes (1981) and Menzel (1974), stated that orientation gestures can be achieved with the whole body in animals, Butterworth argued that both gaze and finger pointing are required for proper indication in human beings (Butterworth, 2003). Therefore, if a robot is to be used adequately as an avatar in a telepresence setting where the intent is for the avatar to indicate objects of interest to co-located individuals, it will require certain body parts namely a head and an arm.

While the taxonomy defined in the previous section classified robots with respect to their use, there is another element with respect to which the robots can be classified and it is with respect to the form they exhibit. One category in this classification is humanoid robots. A robot is usually classified as a humanoid robot when it possesses the following parts, an arm, a torso and a head. While some humanoid robots tend to mirror the human body in terms of both form and size for example the "Actroid DER3" (See Section 3.2.1.3), other humanoid robots are limited to portraying just the top half of a human being for example "BERTI" (See Section 3.2.1.2). Finally, size is a key characteristic of humanoid robots, while some are proportioned to not only mimic a human being in term of form but also in size, others are of a much smaller scale, for example the NAO (See Section 3.2.1.9) or Nexi (See Section 3.2.1.8).

In this section, a number of humanoid robots are going to be listed and described. Although this list cannot hope to be an exhaustive list, it consists of a wide range of well-known and popular platforms that are well suited to displaying gestures.

3.2.1.1 RoboThespian

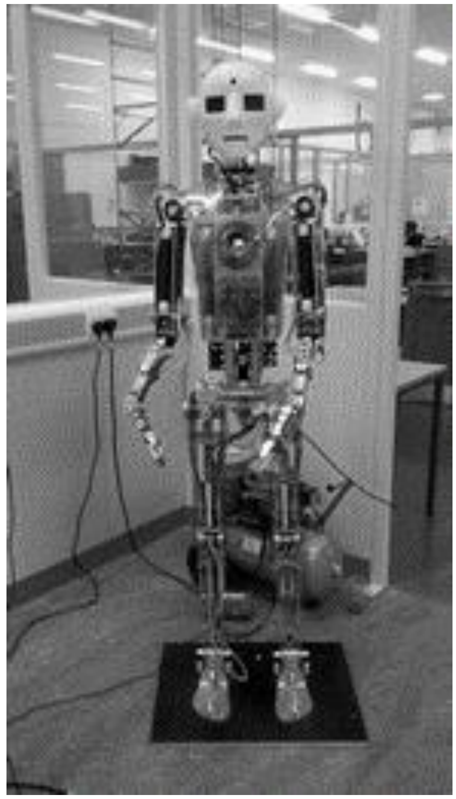

Figure 3.2-A: RoboThespian

RoboThespian™[2] (See Figure 3.2-A), developed by Engineered Arts Limited, is a full body humanoid robot that has been designed to be a robot actor, that is, a robot designed specifically to perform gestures just like actors in plays or movies. The robot has a full body and has a height of 1.7 metres, proportioned similarly to a human, and weighs 30 kg. It is to be noted that while it possesses a full body it is only the top half of the body that is actuated. This is accomplished with

[2] RoboThespian™: ***https://www.engineeredarts.co.uk/robothespian/***

30 degrees of freedom with 10 in each arm and 4 in the head, and the rest in the hips and torso (Engineered Arts Ltd, 2015). It is powered by both pneumatic muscles to allow it to move in a manner similar to a human and make it compliant, as well as by electric motors for degrees of freedom not suited to pneumatic muscle actuation. Despite its size and the high number of degrees of freedom, the cost of the robot is rather low for a human-sized humanoid robot at approximately £30,000.

The physical design of the robot makes it inherently one that is relatively safe. Beyond the pneumatic based actuation mechanism, which allows for some degree of compliance, the light frame of the robot and the limited speed of execution result in a low actuation power. These features are highly beneficial within the context of close interaction with humans; they allow the robot to interact safely, as any possible impact will be of low energy. Additionally, as the robot is human-sized and has human-like features, the use of gestures in conjunction with the robot are facilitated by the body of the robot (Riek, et al., 2010), allowing for gestures to be mapped to the robot frame more simply.

Although some level of compliancy is supported by the use of the pneumatic muscles, the robot itself is not fully compliant. Further, the lack of sensors that could register force being applied to it, mean it is unable to detect collisions with objects or people in the environment. This is an important issue in the context of immediate human interaction, this robot would be ill-suited for such an interaction was it not for the light frame of the robot. The light frame design coupled with pneumatic muscles (providing some compliance) allows the robot to move quite quickly while still not presenting significant danger to objects around it even during collisions. While this is an advantage in terms of safety, it also comes with a drawback as the lack of power due to the light frame and pneumatic muscle combination prevents the robot from doing useful physical work (e.g., manipulating objects).

Another element with respect to safety where human robot interaction is concerned, is the level of safety of the robot. Safety in this context can be of two types, the actual safety level and the perceived safety. For this robot its safety applies to both types. Firstly, the features previously mentioned allow the robot to be perceived as less of a threat and therefore allows the robot to be adequate for the role of embodiment. Secondly, the actual level of safety is also high since any potential collision from the arms produces a maximum dynamic power of 10W, a value obtained from the mass of the arm of 1.5 kg and moving at the highest speed of the robot, which is 0.65 m/s. This maximum amount of power is well below the safety limit imposed by ISO Organisation in their ISO 10218-1 document (ISO, 2006) (dynamic power < 80W) and unlikely to result in any bodily harm (Haddadin, Albu-Schäffer, & Hirzinger, 2009). Moreover, in the case of such a collision, taking into account to the structure of the robot, as described previously, and its compliant nature only a small fraction of that dynamic power would actually be transferred to the object that it made contact with.

Despite the advantages put forward so far with respect to the robot and its possible use in the portrayal of the actions through gestures, it nonetheless lacks certain degrees of freedom and in some cases the full range of motion that can usually be accomplished by a human being. These limitations are design choices and helps to support the level of safety from the robot, while at the same time allowing the robot to perform a wide range of gestures. These in turn prevent the robot from performing certain actions exactly the way humans would carry them out. This means, that while the robot is able to reach a point, the way that it does it may appear awkward. In order to compensate for this an appropriate control module could be implemented, while the control module will not produce gestures that are exactly as a human would perform them, the gestures are expected to be similar enough for the correct meaning to be conveyed. This is supported by the work of Riek et al. (2010), where it was shown that gestures need not be perfectly humanlike to be correctly perceived. Hence this may not be a hindrance to the robot’s utility as an avatar.

3.2.1.2 BERTI

Figure 3.2-B: BertI

The humanoid robotic torso BERTI[3] (See Figure 3.2-B), Bristol Elumotion Robotic Torso 1, is a robot that was developed jointly by the Bristol Robotics Laboratory and Elumotion Ltd, as a multi-purpose robot able to perform a range of activities from pointing, through grasping, to engaging in joint activities with human beings. The robot while being only a torso, arms and head has a height of 1.2 metres and is fixed atop a table 1 metre high. It weighs approximately 60 kg and contains

3 BERTI: ***http://www.brl.ac.uk/***

36 degrees of freedom with 16 in each arm and 2 in the head, and 2 in the hip area. The limb proportions, positions and ranges of motion for the different degrees of freedom are similar to that of human beings. All of the degrees of freedom are actuated by brushless DC motors that allow the movements of the robot to be very fast and precise enough to be able to successfully complete grasping tasks, making it ideal to interact with real world objects. Due to its size and the intricacy of its design, this custom built robot would cost around £300,000 if it was to be obtained commercially.

This design for the robot allows for high speed and accuracy for the execution of actions and general movement but comes at a price with respect to safety. Since the robot is non-compliant, fast moving and has a heavy frame, any collisions are likely to be of high energy. Further, there is an absence of sensors, such as pressure sensors, that could indicate when it has made contact with something in its environment. When the mass of an arm, approximated to about 5 kg, is coupled with a movement velocity of about 1 m/s, a velocity achievable by the robot during normal operation a dynamic power of about 50 W is obtained. While this value is still under the safety limit of the standards put forward by the ISO Organisation in their ISO 10218-1 document (ISO, 2006)(dynamic power < 80W) and not strong enough to result in severe bodily harm (Haddadin, Albu-Schäffer, & Hirzinger, 2009), it is still large enough to severely disrupt the interaction process. This makes it inadequate for use in the immediate vicinity of humans.

The ability of this robot to perform communicative gestures has been previously assessed and used by Riek et al. among others, where the robot was used to portray gestures to participants who were tasked with identifying which gestures was being portrayed by the robot (Riek, et al., 2010). Moreover, the precision in the movement of the robot was also investigated in the work of Broun, Beck, Pipe, Mirmehdi, and Melhuish (2014).

3.2.1.3 Actroid DER3

Figure 3.2-C: Actroid DER3

The Actroid DER3[4] (See Figure 3.2-C), is the latest model in the Actroid series a type of humanoid robot that was originally developed by Osaka University. The robots in this series do not have any mobility functions, while they can move their legs they cannot use them to walk. Being immobile robots, they are usually used as guides or presenters on top of a presentation platform at exhibitions. Moreover, the finish of the Actroid series makes it reminiscent to an adult female human being of Japanese descent. The Actroid DER3 has a height of about 1.8 metres and weighs approximately 100 kg. It is composed of 55 degrees of freedom allowing for not only the actuations of fingers, arms, torso and head (as in the previous examples) but also allow some leg control and

[4] Actroid DER3: ***http://www.kokoro-dreams.co.jp/english/rt_tokutyu/actroid.html***

actuation of facial features such as the eyes, cheeks and lips (Kokoro Dreams, 2015). Contrary to the previous humanoids, the DER3 can only be obtained on a rental system at about £500 per day for customers inside Japan.

The Actroid DER3 robots share a number of characteristics with the RoboThespian, these not only include the structure of the body and the size of the robot but also the method of actuation. The DER3 robots, similarly to the RoboThespian, possess pneumatic muscles (Kokoro Dreams, 2015) and use them to control the gestures that the robot can perform. Contrary to the RoboThespian design however, the design of DER3 does not only possess the features that make it human-like as the RoboThespian, but the resemblance to actual human beings is higher. This close resemblance to an actual human being can provide the robot with some advantages in certain circumstances such as interacting with objects designed for humans. However, in the case where such a robot is to interact with human beings there is one important factor to take into account. In section 2.2.3 while discussing the form of humanoid robots, it was stated that there was a phenomenon called the “uncanny valley” that affects their acceptance. As mentioned earlier the term was first coined by Mori (1970) and further extended in the work MacDorman and Minato (2005). According to this line of reasoning, an appearance that is too human-like can raise a sense of unease in the person interacting with such an object. Considering that the Actroid DER3 is very human-like and meant to portray the actions of a human being, it is likely to be affected by this phenomenon. Furthermore, using such a robot for tele-presence can lead to confusion, especially in the case where the one being represented does not resemble the robot.

3.2.1.4 ASIMO

Figure 3.2-D: Asimo

The ASIMO[5] (See Figure 3.2-D), Advanced Step in Innovative MObility, is a robot that was developed by Honda. It was developed as a walking robot platform that would also be able to provide assistance to humans in various situations (Sakagami, et al., 2002). Being designed to walk and provide assistance, the ASIMO robot possesses two walking legs and also two arms with which it can perform activities ranging from pointing, to grasping and lifting, and also carrying

---

5 ASIMO: ***http://asimo.honda.com/***

the object to a destination. It is also move through an environment designed for humans and is able to use stairs. This robot is an ongoing project and therefore there are different iterations of this robot. The current one was designed in 2011. The robot has a full body with most of the joints in a human body mirrored in the robot. It is 1.3 metres in height and weight at about 54 kg. It also contains 57 degrees of freedom with 20 in each arm and 3 in the head, 2 in the torso and finally 6 in each leg (Honda, 2015). Similarly to the BERTI robot the degrees of freedom are implemented using DC motors which allows the robot to carry objects up to 1kg, though the weight of the object carried impacts the maximum achievable speed of the robot (Honda, 2015). Using its legs the unloaded robot is able to reach a maximum speed of 1.67 m/s (Honda, 2015). Since it is not a commercially available robot, it is difficult to put an actual price to the robot but it has been estimated to cost about £630,000.

The ASIMO is a very advanced humanoid robot with a number of sensors that can detect objects around it and also can handle a number of navigation tasks such as climbing up and down stairs and also avoid obstacles present in its path (Honda, 2015). Due to its versatility, this robot has been used in a number of projects both in terms of bipedal navigation research as in the work of Chestnutt et al. (2005) and also in terms of human robot interaction as in the work of Mutlu, Osman, Forlizzi, and Hodgins (2006) and Mutlu, Forlizzi, Hodgins, and Kiesler (2006).

3.2.1.5 iCub

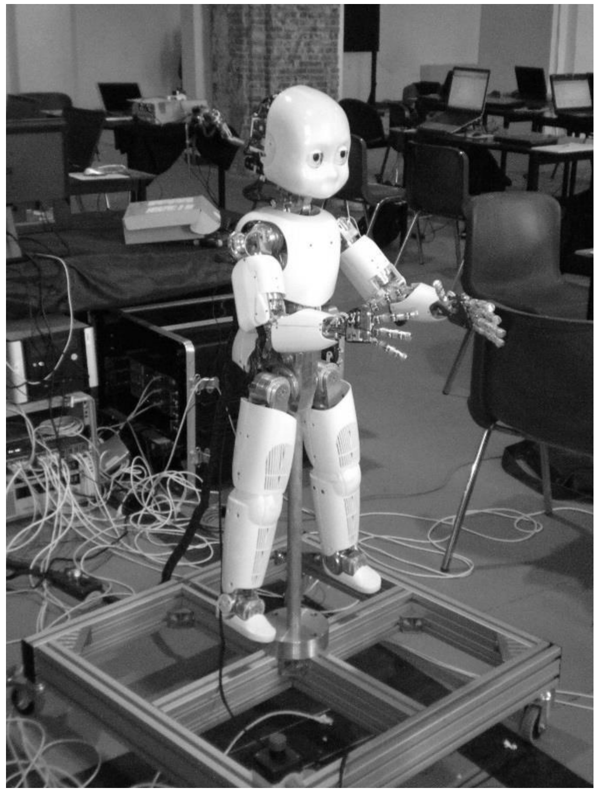

Figure 3.2-E: iCub

The iCub[6] (See Figure 3.2-E), is a robot that was developed by the Italian Institute of Technology (IIT). It started as an offshoot of an EU project, RobotCub, aimed at studying cognition with humanoid robots. The iCub as its name implies is meant to portray a child and it has a height of 1 metre and a weight of 22 kg. The robot contains 53 degrees of freedom with 16 in each arm and 6 in the head, 3 in the waist and 6 in each legs (Tsagarakis, et al., 2007). As in the previous robot,

---

6 iCub: ***http:// http://www.icub.org/***

these degrees of freedom are implemented using DC motors that allows for precise movements. The robot is able to grasp small and light objects and also coordinate both its arms to accomplish two handed tasks. It is able to move around using two modes of travel, it can either walk upright on two legs or crawl on all fours. Its maximum achievable speed while on two legs is 0.7 m/s (Lu, Lallee, Tikhanoff, & Dominey, 2012).

While the iCubs are no longer commercially available, they are more common than the ASIMO robots with about 30 iCubs spread across the world and are owned by various universities and other research institutes. When they were commercially available, the price of obtaining one was about £200,000.

As detailed above, the iCub has been designed as a child and therefore has a childlike appearance and dimensions. These robots have been used in a number of situations that would be appropriate to an infant, such as projects where there were word object pair association (Metta, Sandini, Vernon, Natale, & Nori, 2008) or other simple cognitive tasks (Tsagarakis, et al., 2007). Its supported childlike mobility was also been used in certain projects to find the proper way to move using this robot (Degallier, et al., 2008).

3.2.1.6 Wakamaru

Figure 3.2-F: Wakamaru

The Wakamaru[7] (See Figure 3.2-F), is a robot that was developed by Mitsubishi Heavy Industries This robot was developed as a commercial product that would assist and provide companionship in domestic situation. The robot is 1 metre tall and weighs 30 kg. Contrary to the previous robots, this robot uses wheels for mobility instead of legs. The robot contains only 13 degrees of freedom

[7] Wakamaru. ***https://www.mhi-global.com/products/detail/wakamaru_spec.html***

with 4 in each arm and 3 in the neck, and 2 for the wheels (Namera, Takasugi, Takano, Yamamoto, & Miyake, 2008).

As with the previous robots the degrees of freedom are implemented using DC motors. The limitation in the number of degrees of freedom, prevents the robot from interacting directly with objects or limits the amount of gesturing. The type of interaction preferred by this robot is audio which allows it to first identify an interlocutor and then engage with him verbally (Shiotani, et al., 2006). Being designed for a domestic environment, the mobile platform contains wheels and proximity sensors to prevent it bumping into objects. The mobile platform allows it to move at a maximum of 0.3m/s on flat surfaces.

Since the Wakamaru was designed to be a commercial robot for domestic settings, the price of the Wakamaru is considerably lower than the other robots, at only about £9,000 per unit.

As can be seen from the associated picture, the Wakamaru while having hands, the hands themselves are under actuated, to the point where the gestures that can be done with the robot are limited to the use of the arms. Nevertheless, it has been used in a number of projects as a companion / support robot in home settings as in the work of Shiotani et al. (2006).

3.2.1.7 Robovie R3

Figure 3.2-G: Robovie R3

The Robovie R3[8] (See Figure 3.2-G), is a robot that was developed by Vstone Co. Ltd. Similar to the Wakamaru robot, the Robovie R3 robot was also developed as a social robot. To that effect it is the social aspect of interaction that is focussed on in the robot design, consequently it is able to perform some gestures but is unable to lift anything but light objects.

8 Robovie R3. ***http://www.vstone.co.jp/english/products/robovie_x/***

The robot is about 1 metre tall and has a weight of about 35 kg. Similar to the Wakamaru robot it is a robot torso on top of a wheeled platform. The robot contains 17 degrees of freedom, implemented using servomotors, with 4 in each arm, and the others present in the head and neck and wheeled platform (VStone, 2015).

Similar to the Wakamaru it has a fixed faced devoid of expression, it is able to use its arms to gesture in a limited capacity using some arm gestures. Unlike the Wakamaru it is able to grasp light objects and can point to objects using its arms. It is able to move quite rapidly with respect to its size achieving a maximum speed of 0.7 m/s (VStone, 2015). However, for its size and specifications it is rather expensive especially when compared to the Wakamaru robot as the price is almost tripled with the basic unit available for about £27000.

As mentioned previously this robot is a commercial product with similar specifications to the Wakamaru robot and similar to the Wakamaru robot, it has under actuated hands, while these hands are able to hold light objects, they do not have enough dexterity for performing meaningful gestures.

3.2.1.8 Nexi

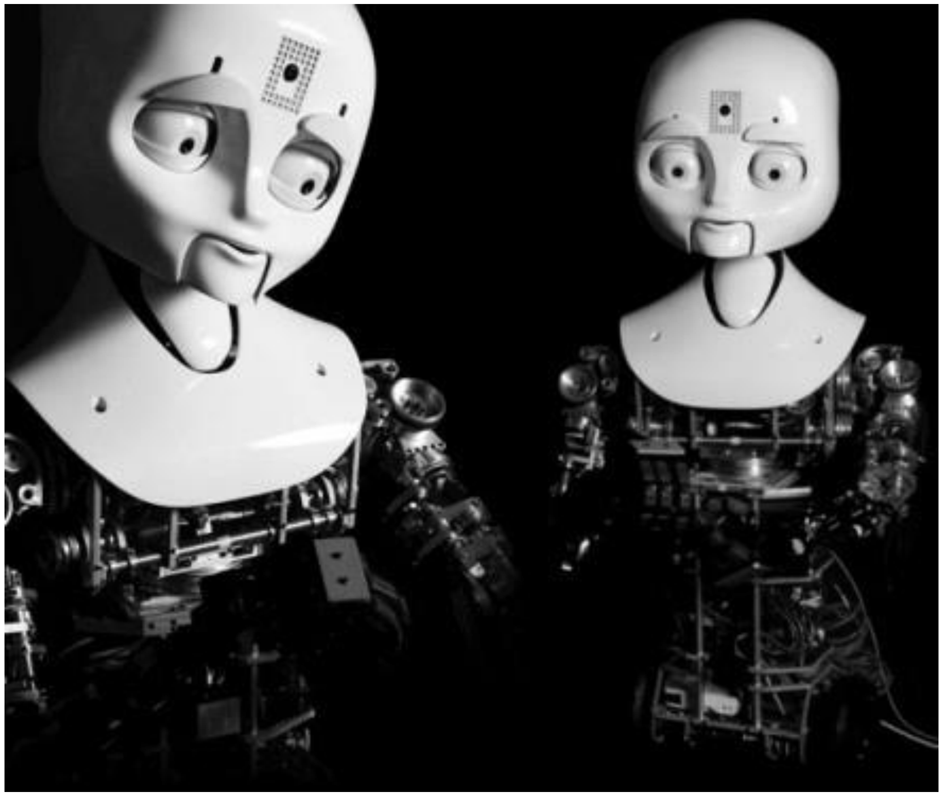

Figure 3.2-H: Nexi

The Nexi[9] (See Figure 3.2-H), is a mobile humanoid robot that was developed by the Massachusetts Institute of Technology. While it shares a lot of features with the Wakamaru, the Nexi is different from the Wakamaru in that it is a robot developed solely for research purposes and is focussed around expressiveness rather than assisting.

[9] Nexi. ***http://robotic.media.mit.edu/***

The robot is 0.9 metre tall and weighs 14 kg. Similar to the Wakamaru, the Nexi also uses wheels for mobility instead of legs. The robot contains 34 degrees of freedom with 8 in each arm, 16 in the head, and 2 for the wheels (Xitome Design, 2015).These degrees of freedom are only implemented using DC motors.

Contrary to the Wakamaru robot this robot is able to grasp and move moderately heavy objects of up to 5 kg (Beciri, 2015). The robot can use both its arms and its face to interact with an interlocutor. The range of facial expressions that can be achieved by the robot together with the use of arm gestures allows the robot to operate in a social setting. While it is a mobile robot like the Wakamaru robot, it is limited to a maximum speed of 0.15 m/s (Xitome Design, 2015).

Since it was developed as a research platform and not as a commercial robot it is difficult to put a price on the actual robot but the price of this robot would be about £10,000 (Xitome Design, 2015).

A feature that uniquely identifies this particular type of robot from the others present in this list is the presence of an expressive face. The face in this robot is able to convey a degree of emotions, while at the same time avoiding the face looking too humanlike, which is a useful feature as it will not suffer from the risk of confusion previously mentioned. Using this feature together with its arms, the robot is able to convey a level of presence which can be beneficial in group collaboration settings as shown in the work of Shah, Wiken, Williams, and Breazeal (2011).

3.2.1.9 Nao

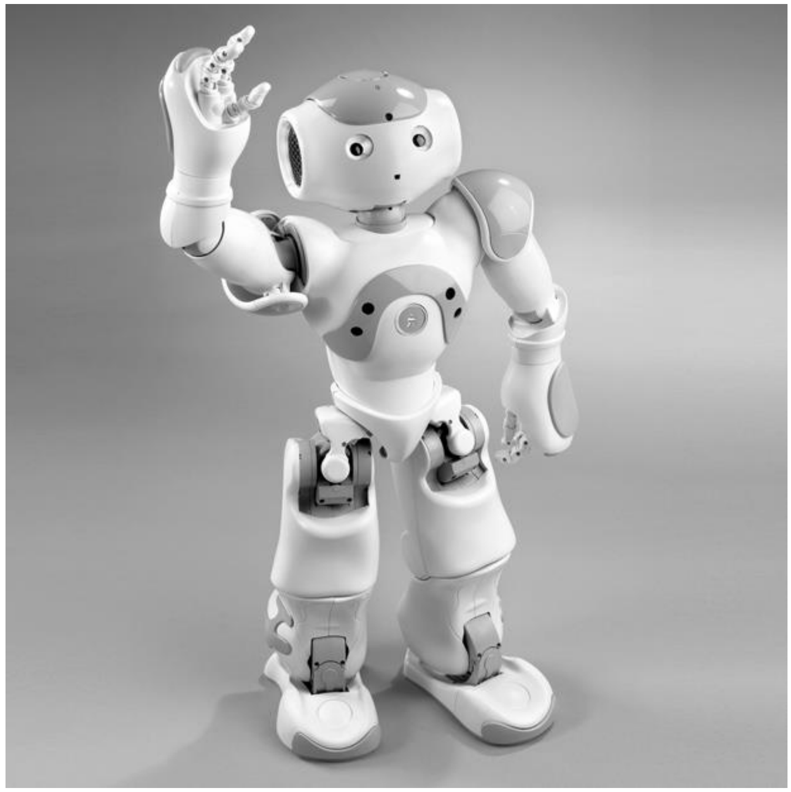

Figure 3.2-I: Nao

The Nao[10] (See Figure 3.2-I), is a robot that was developed by Aldebaran Robotics part of the SoftBank group. It was first developed as a robot platform for robot football competitions, since then it was redesigned and repurposed as a research platform and an entertainment robot.

---

[10] Nao. ***http://www.aldebaran.com/en/humanoid-robot/nao-robot***

The robot is 0.58 metre tall and has a weight of 4.3 kg. Being a full body humanoid robot the Nao is quite similar to the ASIMO though much smaller in size. The robot contains 25 degrees of freedom, implemented using DC motors, with 6 in each arm, 2 in the head, 1 in the hip, and 5 in each leg (Aldebaran Robotics, 2014).

While the Nao has a fixed faced expression, it is able to use its arms to gesture in a limited capacity using arm gestures rather than hand gestures. The Nao is also able to grasp light objects and carry them around. When it is not carrying any object it is able to achieve a maximum speed of about 2 m/s (Gouaillier, et al., 2008).

The robot has been primarily used in a number of research projects ranging from mobility to object manipulation. Compared to the other robots, it has a relatively low price with the basic unit available for about £5,000.

Due to its size and availability, it has been used in a number of research projects, from bipedal locomotion (Strom, Slavov, & Chown, 2010), object identification and tracking (Sauppé & Mutlu, 2014), to even object designation (Sauppé & Mutlu, 2014) and general human interaction (Jokinen & Wilcock, 2014).

3.2.1.10 Comparison of the Robots

In this section, a number of humanoid robots were listed and described. While this list is not an exhaustive list of the humanoid robots, it lists some of the most appropriate robot for the requirements of the task that the humanoid robot is to perform. As mentioned before, some robots such as the Actroid series, while being humanoid in appearance are unfit for use in this thesis as they have a humanlike face and this could prove problematic if they are to act as embodiment for a remote person. From the list of features listed for each robot in this section, a table of specifications comparing the different robots has been created (see Table 3.2-I) to aid comparison. The dimensions used for the comparison between the different robots are with respect to the task that the robots are to be used for in this thesis and also with respect to their availability.

Table 3.2-I: Humanoid Robots Basic Specifications

Name	**Robo-Thespian**	**BERTI**	**Actroid DER 3**	**ASIMO**	**iCub**	**Wakamaru**	**Robovie R3**	**Nexi**	**Nao**
Height (m)	1.7	1.2	1.8	1.3	1	1	1	0.9	0.58
DOF(Usable)	26 (25)	18 (18)	55 (18)	57 (45)	53 (39)	13 (11)	17 (7)	34 (19)	25 (13)
Price	~£30,000	~£300,000	~£500 / day	N/A (~£630,000)	~£200,000	~£9,000	~£27,000	N/A (~£10,000)	~£5,000
Actuation	DC Motor, Pneumatic Muscles	DC Motor	Pneumatic Muscles	DC Motor	DC Motor	DC Motor	Servo Motor	DC Motor	DC Motor
Face	Fixed	Blank	Expressive	Blank	Fixed	Fixed	Fixed	Expressive	Fixed
Hands	Individual actuated fingers	Individual actuated fingers	Individual actuated fingers	Individual actuated fingers	Individual actuated fingers	Unactuated fingers	Unactuated fingers	Individual actuated fingers	Grouped actuated fingers

3.2.1.11 Selection of the humanoid robot

The humanoid robotic platform that was selected in this thesis is the RoboThespian, it has a number of advantages over the other robot platforms described above for use as an avatar. First of all, the robot has a full humanoid body which includes a face and eyes with which a form of gaze can be created. It also has long arms together with individually actuated fingers, which can be used to create clear pointing gestures and other hand gestures. One important element that makes this robot useful as an avatar that is to be used in proximity to human beings, is its safety features. The first of these is its light frame, which tends to minimise the damage of any impact. This is coupled with a form of quasi-compliance. While the robot lacks specific sensors that would enable active compliance, its design allows for a degree of physical manipulation, enabling a person to move it slightly out of the way. This compliance is obtained from the pneumatic muscles actuators, often coupled with elastic elements in the robot. Although the design of the robot precludes useful physical work (such as manipulation of objects) this is not required in the context of this thesis.

Another important element in the decision to use the RoboThespian is the availability of the robot in the location where the research was carried out. The Actroid DER3 has a number of similar characteristics to the RoboThespian but its limited geographical distribution and the presence of a human-like face, which can be affected by the uncanny valley, as mentioned previously, were factors that worked against its selection. Another suitable candidate for the function of the humanoid robot in this thesis is the Nexi, but yet again the limitation in terms of availability, limited to a fixed research group, prevented its use. Finally developing a custom humanoid robot, or indeed reproducing one of the custom robots described above, was not considered as it is beyond the scope of this thesis.

While RoboThespian proves quite useful for the required tasks, there are still a few limitations that impact its usefulness. The restricted motion in terms of range and speed coupled with the lack of some degrees of freedom are examples of such limitations. The same reasons for why it is good from a safety point of view, for human interaction, also become drawbacks when it is required to perform gestures captured from human motion. This, however, presented an opportunity to explore the perception of motion that would usually just be assumed, resulting in a method for producing useful gestures while remaining sage for human interaction. This opportunity and a solution to the above problem came in the form of a custom-made control module.

More information on the control module together with its submodules will be provided in later chapters. However, since the term control module can be used, with respect to robots, to describe any behaviour that it enables or monitors on the robot, this specific control module is specifically linked with gesture generation for the robot. Consequently, this module involves both the enabling of gestures for the robot and the verification that the enabled gestures are human-like. The algorithms used for these tasks will be described and explained in the following chapters.

3.2.2 Non-humanoid Robots

While in the previous section the focus was on humanoid robots, in this section the focus is on non-humanoid robots. Non-humanoid robots are robots that do not possess the features that will qualify them as having a humanoid form. There is a great number of non-humanoid robots in use, a lot of them as industrial robots used in different sectors. Since industrial robots do not have to operate in environments primarily designed for humans the form of such robots can be reduced in complexity to just the components required for such tasks.

The non-humanoid robots that are described in this section are limited to a list of the non-humanoid robots that were used in the experiments carried out as part of this thesis. They were selected to provide a contrast to the humanoid robot that was selected as the humanoid robot platform. Since the task that the embodiment has to accomplish varies between experiments two non-humanoid robots were required to provide an appropriate counterpart to the humanoid robot.

3.2.2.1 ABB IRB 120

Figure 3.2-J: ABB IRB 120 Arm

The first of these non-humanoid robots is the ABB IRB 120[11] which will be referred to as the ABB Arm in this thesis (See Figure 3.2-J). The ABB arm is a product of ABB. It is an industrial robot arm designed for small-scale manufacturing. As is to be expected from an industrial robot, not only is the form of the robot completely different from that of the humanoid robot but also its basic design is markedly different. Industrial robots are designed for use in industrial settings, they are therefore able to perform repetitive high precision tasks at high speed. The aesthetics of this robot reflect the intended application. It is made up of cylindrical or cuboid shaped components assembled to form of a multi-jointed arm.

---

11 ABB IRB 120. ***http://www.abb.co.uk/***

The ABB arm, is a light industrial robot, weighing 25 kg with an effective maximum load support of 3 kg. It has a wide range of motion along each of its 6 degree of freedom, able to describe an arc of about 330° at its base and with its minimum joint range being 180°. As mentioned previously, industrial robots are highly accurate specially at repeating the same task and in the case of this robot having a repetition accuracy of 0.01 mm. It is also able to move each joint at a maximum angular velocity of 250°/s, or 0.7 rotations per second. This results in a net dynamic power of about 75W with a 3 kg load attached, which is just under the safety limit (dynamic power < 80W) set by the ISO Organisation in their ISO 10218-1 report (ISO, 2006).

Despite this robot being a light industrial robot and within the safety limit, the impact from this arm making contact at full speed with a person can still cause injury, especially since it has no compliance by which to lower the impact energy. So the use of industrial robots in general and this robot in particular, in environments where humans are present in the immediate vicinity are usually uncommon. While not ideal for performing human-like gestures, it was used in a limited capacity in this thesis, due to both safety concerns and lack of essential elements for proper gesture portrayal.

Since this robot is both accurate and can move quite quickly, it is a suitable candidate for being used in a tele-pointer capacity and as such can be used to provide a stark contrast to the RoboThespian. Since it was designed for use in small-scale manufacturing, it has a lower load capacity than other industrial robots. However, this is not of great importance for this thesis.

3.2.2.2 Cyton Gamma 1500

Figure 3.2-K: Cyton Gamma 1500

Another non-humanoid robot that was used in the investigation was a Cyton Gamma 1500[12], a product of Robai, which will be referred to as the Cyton arm in this thesis (See Figure 3.2-K). This robot looks similar to an industrial robot although it is less bulky and less strong. It is a very light robot weighing 2kg, containing 7 degrees of freedom and also possessing a two finger gripper. The ranges of motion of the joints are limited to about 180^{o} for each joint. It is able to operate with 1 mm precision. Relative to the ABB arm, it is slow, with its joints capable of an angular velocity of 90^{o}/s. It is able to lift a 1.5kg payload.

Being significantly lighter and moving more slowly than even a lightweight industrial robot, this robot provides a higher level of safety, an essential element for operating in close vicinity to human

---

[12] Cyton Gamma 1500. ***http://robai.com/robots/gamma-1500/***

beings. However, it is to be noted that it is not compliant. In this thesis, this robot is used in situations where a remotely located human being interact with other humans, via the non-humanoid robot, using gestures. This is facilitated with having many degrees of freedom with which the gestures can be adequately portrayed. Even the end-effector, which is a gripper, is used as part of the gesture portrayal function though the ability for it to grasp is incidental to its use in this thesis.

3.2.2.3 Summary

In the previous section, a range of humanoid robots were described, and one in particular was selected based on the requirements of a robot avatar that can perform human-like arm gestures. A non-humanoid robot will also be used, in some experiments, to investigate the benefits, if any, of the humanoid form for robotic embodiments. The two robots, described in this section, are those alternative robots that are going to be used to contrast with the humanoid one. Each of the robots, in this section, has a specific purpose and reasons for being selected for the comparison process. Firstly, the ABB arm is a strong and very accurate robot able to point precisely and concisely to a specific point making it the ideal tele-pointer, as described in Chapter 2. It will therefore be used in Chapter 4 and Chapter 5, as an alternative to the RoboThespian to help in identifying benefits of the humanoid form for such tasks. On the other hand, the Cyton arm is a light and accurate robot which possesses specific features that allows it to better perform certain gestures, and operate in the same workspace as a person. This makes this robot a good choice as an alternative to the gesture and social tele-presence aspect of this thesis, which are going to be described in Chapter 7. These robots while being different, with specific abilities and limitations, serve a similar purpose in this thesis. This purpose is to act as a non-humanoid alternative to evaluate the advantages of the humanoid form for these tasks.

3.3 Tabletop Interactive Devices

In the setups that were investigated in the previous chapter, there was a lot of emphasis on duplicating the experience of the tabletop interaction at each location. One of the reasons for this is to create a shared frame of reference when addressing participants about objects (Gaver, 1992). It was therefore critical for the experiments carried out that a similar setup was provided at the different locations in order to minimize errors arising from orientation issues such as those mentioned in (Hindmarsh, Fraser, Heath, Benford, & Greenhalgh, 1998).

3.3.1 Microsoft Surface

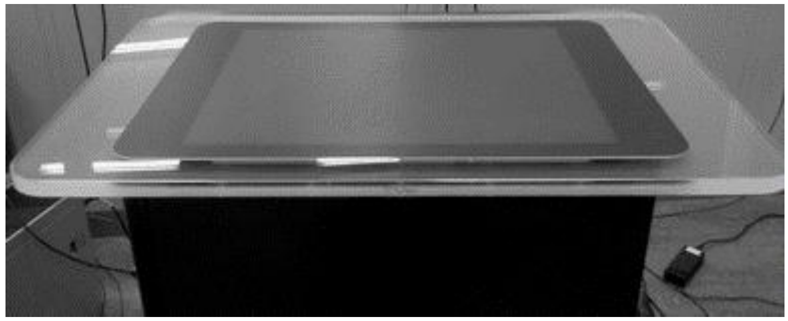

Figure 3.3-A: Microsoft Surface

The tabletop interactive device (TID) that was used in this thesis is the Microsoft Surface®[13] V1.0 (See Figure 3.3-A). As with other TIDs, it is a multi-touch screen that allows people around it to interact with a program using touch technology. The screen makes use of a projector with a resolution of 1024 by 768 pixels and is 625 by 460 mm in dimensions. The Microsoft Surface uses infrared technology in order to detect objects that are either touching its top directly or hovering just above. When it is used with its software development kit (SDK) and in combination with Microsoft XNA technology it allows the users to touch, drag or otherwise manipulate virtual objects by uniquely identifying each contact with its surface. Each of these contacts needs to have a diameter of at least 8mm for them to be recognised by the software of the table.

[13] Microsoft Surface® is a Product of Microsoft®. ***http://www.surface.com***

3.3.2 Horizontal Screen

Figure 3.3-B: Horizontal Screen

In order to duplicate the presence of the Microsoft Surface® in the environment of the remote user, one possible option was to use another such device. However, the acquisition of another such device, apart from being costly, could have interfered with a component from the remote input capture system, the body pose capture device. This is because both of these devices are based on a similar technology, therefore an alternative option making use of a horizontal screen (See Figure 3.3-B) was used. The screen here is a flat screen that has been made to lie horizontally using a custom built frame. While the screen itself is not interactive, as it has no sensors with which it can read the interaction that the remote user is performing, it provides a large enough area. This allows the remote user to perceive a same scale rendition of the interaction happening on the Microsoft Surface®; the scaling is essential for the common frame of reference, previously introduced. In order to properly duplicate the interactive experience of the Microsoft Surface®, a screen with a minimum displayable area of 625 by 460 mm was required in order to accurately represent the interaction of the other user. The body pose capture device was used to make the screen behave like a TID, the steps do so are described in section 6.2.1.

3.4 Software Components

In the previous sections of this chapter, the focus has been on the hardware components, ranging from the main components such as the robot and tabletop interactive devices (TIDs) to the communication peripherals. However, these hardware components were unusable without accompanying software modules. In the case of the robots used in this thesis, the custom control modules were implemented to allow them to perform task in a particular way. For the TIDs, custom network enabled software applications, using a client-server architecture, were required in order to allow the same task to be visible to the different groups of people. Apart from these tailor-made software modules, off-the-shelf applications were also used in order to enable audio communication between the different groups. More information on these software components will be provided in the specific chapters where they were used.

3.5 Chapter Conclusion

The purpose of this chapter was to list and describe the main components, from both hardware and software, that were used in the experiments. Due to the nature of the topic investigated, that is Robot Mediated Communication: Enhancing Tele-presence using an Avatar, there were a number of critical hardware components required for the research. These hardware components are namely a robot platform, to provide a presence, and a tabletop interactive device (TID), to provide a suitable interactive environment. While the robot platform was used in order to allow a remote presence to be emulated it is through the use of TIDs that the individuals comprising the group can interact with the others. Except for those critical components, a number of secondary components were also required for properly conducting the experiments; those components being specific to the actual experiments are going to be described in the relevant chapters.

Chapter 4 Robot Pointing

4.1 Chapter Introduction

In the previous chapter, a description of the main hardware components together with their supporting software used in the investigative part of this thesis was provided. As part of the description their capabilities and limitations were stated as well as the reasoning behind them being selected for the investigation. Since the focus of this thesis is on Robot Mediated Communication, as a first step in the investigation, it was important to figure out what the impact of the use of a robot was on basic interaction. Consequently, in this part of the investigation, a simple pointing task was implemented, using robots as the indicators. It is to be noted that while the overall idea behind this investigation is the use robots to facilitate interpersonal communication between groups of people over vast distances, this chapter uses robots as autonomous agents, in the sense that an individual is not dictating to the robots where to point to.

In this chapter, this basic pointing task is implemented using two robots, one humanoid and the other an industrial robot arm. This task is different from the pointing gesture previously described in Chapter 2, in that, it only consists of the body movement involved in pointing. This is different from the context that is associated with pointing, which is usually part of a discourse, where the gesture is accompanied by words. As part of the task, each robot is used to point to different locations. The subject then has to identify which location the robot is attempting to indicate, which is then recorded and analysed, to find out how successful the robots are at communicating through pointing alone.

4.1.1 Regarding Pointing

In Chapter 2, the general description provided for gestures included both their importance as part of discourse, and the different classification of gestures. One of these classes of gestures is deictic gestures. This class of gestures is used to refer to physical elements that are being used as part of the speech. One subclass of these deictic gestures is the pointing gestures. While pointing gestures are usually carried out through the use of the finger pointing gesture (Kita, 2003), it can also be carried out in a number of ways (Goodwin, 2003). On the topic of finger pointing gestures, it should be noted that finger pointing is a uniquely human gesture (Butterworth, 2003). While the actual gesture is not truly universal, since some cultures use a different configuration of the fingers (Wilkins, 2003), the accepted configuration of the hand for finger pointing gesture, with the index finger being straight while the others are closed, can be used in different ways (Kendon & Laura, 2003). They can be used to uniquely identify specific objects or groups of objects depending on the context (Kendon & Laura, 2003). This type of hand gesture is usually supported by the use of the gaze, that is the interlocutor uses the direction to where he is looking to help quickly shift attention to the area being designated (Kita, 2003). While the gaze provides a general direction it is not specific enough for accessing the real direction due to the size of the lever (Butterworth & Itakura, 2000), that is the distance from the fulcrum is too small for the observer to uniquely identify which is the object that is being looked at naturally. The importance and popularity of the pointing gesture using both finger pointing and gaze makes it an important element to implement in a tele-presence robot. In a tele-presence setting pointing gestures are particularly essential since current standard in peer-to-peer video communications, is often not sufficient to show the full range of gestural motions in space (Kauff & Schreer, 2002) and a robot would be able to mitigate this problem especially for listeners when co-occurring speech is ambiguous (Lemme, Freire, Barreto, & Steil, 2013).

One element of implementing pointing in a robot is the type of pointing to implement. Pointing via artificial bodies have been implemented with varying success in a number of ways (Kuzuoka, Oyama, Yamazaki, Suzuki, & Mitsuishi, 2000; Hindmarsh, Fraser, Heath, Benford, & Greenhalgh, 2000), and this includes both real world robots (Kuzuoka, Oyama, Yamazaki, Suzuki, & Mitsuishi, 2000) and virtual avatars (Hindmarsh, Fraser, Heath, Benford, & Greenhalgh, 1998; Hindmarsh, Fraser, Heath, Benford, & Greenhalgh, 2000). Some implementations of the pointing gestures used some additional hardware in conjunction with the actual robot body to decrease the ambiguity in the identification of the object (Kuzuoka, Oyama, Yamazaki, Suzuki, & Mitsuishi, 2000). Implementing a pointing gesture for a robot comprises of two separate issues, the first is the form of gesture being implemented and the second is related to the purpose of the gesture. Regarding the former issue, considering that robots come in different forms and sizes and even in humanoid robots a five fingered hand is not always implemented (Namera, Takasugi, Takano, Yamamoto, & Miyake, 2008), a direct copying of the index pointing gesture is not always possible. However, while pointing using the index is uniquely human (Butterworth, 2003), animals ranging from dogs to primate have been taught to point to indicate objects of interest (Hewes, 1981). While using the primate pointing paradigm involves the use of arm, the canine pointing paradigm on robots would make the robot point with its whole body (Hewes, 1981). One added incentive of using such specifics pointing models is that according to Oberman et al. comprehensive robotic gestures might activate the mirror neuron system that was previously thought to be specifically selective for biological actions and thus allow the gestures to be more meaningful (Oberman, Pineda, & Ramachandran, 2007).

On the second issue which relates to the purpose of the gesture, like other gestures, pointing gestures are used to facilitate language processing in humans (Butterworth, 2003; Freedman, 1977) but in its own unique way. While other gestures tend to have only a form that the interlocutor has to correctly identify in order to understand what meaning is being conveyed, pointing is usually used in conjunction with the environment, both the external one and the internal representation of it in the interlocutors' representation (Goodwin, 2003). So while the person using the pointing gesture can point to real world visible objects (Goodwin, 2003), pointing can also be used to point to objects that are not visible in the real world an example of which being giving verbal directions. However, this necessitates that the pointing be used in conjunction with speech. In the pointing being implemented here, only a body performing a visible act of pointing is required, which is important if a robot is to use the natural communication modalities that are used in Human-Human Interaction (Breazeal, 2003; Goodwin, 2003).

4.1.2 Hypotheses

Subsidiary to the main hypotheses described in Section 2.4.2 a number of sub-hypotheses relating to robot pointing have been devised. As mentioned earlier in the chapter, when pointing is performed by a human being it does not only involve the finger (Goodwin, 2003). While the finger is used to ultimately designate the target object, it is a small lever and other parts of the body, which provide a bigger lever, such as the forearm, also form part of the pointing gesture (Kita, 2003). One element that should also be taken into account is human gaze that can be used to further support pointing by drawing more attention to a specific area or object (Butterworth, 2003). Since pointing is an important element of interaction, the efficacy of a robotic embodiment for pointing needs to be assessed. This can be measured using the speed and accuracy with which indications are identified. This results in the following sub-hypotheses:

H 4.1 A physical embodiment of the action of a remote user will allow for the actions to be identified faster than when no physical embodiment is used.

H 4.2 A physical embodiment of the action of a remote user will allow for the actions to be identified as accurately as when the actions are performed by a co-located human being.

While a local physical embodiment of the pointing from a remote person, could be useful the form of the embodiment and the method used to accomplish the pointing may also play a part in the identification of the target. In the work of Kuzuoka et al. (2000), a non-humanoid robot was used to point to objects in the environment and proved unsuccessful until a laser pointer was used. More recently, in the work of Lemme, Freire, Barreto, and Steil (2013), and Sauppé and Mutlu (2014) humanoid robots have been used to successfully point to objects. However, the form of the robot alone is not the only element to consider when dealing with humanoid robots. As pointed out by Oberman et al. (2007), robotic gestures can trigger the mirror neuron system responsible for the comprehension of actions, similar to the actions carried out by biological entities. Therefore, actions may be more successfully recognised if performed in a suitably biological way. However, in the work of Riek et al. it has been shown that humanoid actions do not have to be exactly human-like to be useful (Riek, et al., 2010). This is important considering that the humanoid robot used in this thesis do not have full human range of motion. Using a control scheme that allows human-like gestures to be performed together with the humanoid robot, this is probably going to be true here. This results in the following sub-hypothesis:

H 4.3 The humanoid robot, with an implemented joint coordination scheme, will be more effective at conveying the actions of a remote user than a standard one for example an industrial robot arm (that is, faster and more accurate identification).

4.2 Method

In order to test hypothesis H 4.3, an appropriate control scheme that produces human-like motion, was required. For that control scheme to be successfully implemented, it was critical to understand how human beings behave, specifically when they are pointing to elements in an environment. Therefore prior to conducting the main experiment to test the suitability of the gestures produced by the control scheme, a preliminary study was first conducted. The preliminary study was carried out to gather data on how pointing is carried out by human beings and thus provides the basis for a pointing scheme that produces humanlike finger pointing. The main study was then conducted to evaluate the implemented scheme when compared to a normal pointing scheme from a non-humanoid robot.

4.2.1 Participants

Six participants took part in the preliminary study (3 females, 3 males); aged 24 - 31 (M = 29.1, SD = 2.56). These six participants were divided into two groups of three members each.

For the main study, a total of twelve participants took part in the experiment (4 females, 8 males); aged 22 - 38 (M = 28.1, SD = 3.08). Each of the participants were asked to identify 4 squares selected at random, for each of the conditions of the experiment.

4.2.2 Material

In order to test the above hypotheses, two different robots were used: RoboThespian, a humanoid robot, and ABB arm, a non-humanoid industrial robot. Descriptions of these robots together with their general capabilities and limitations can be found in Chapter 3.

4.2.2.1 Preliminary Study

In order to structure the environment and provide fixed points which a participant can point to, an interface was implemented on the tabletop interactive device (TID), Microsoft Surface, displaying a number of two dimensional artefacts, for the participants to be able to point to. More specifically, the interface that was implemented presented the participants with a simple jigsaw puzzle that they were tasked with solving. This required the participants to select the different pieces of the puzzle by touching them and dragging them across the surface.

During the study, it was observed that while touch was the primary way in which the participants interacted with the interface to move the artefacts around as instructed; pointing was also used to indicate other artefacts of interest. It was noted that while the pointing was taking place, a number of different poses were used for the same task. Furthermore, the arm poses were observed to be dependent on the proximity of the artefact to be pointed to, and this was implemented in the control scheme to increase the human-likeness of the motion of the robot.

4.2.2.2 RoboThespian (Humanoid Robot)

The limitations presented previously, in Chapter 3, prevent some gestures that are possible for a human to be directly copied by the robot. Of particular importance, is the limited humeral rotation toward the centre line of the torso; this prevents the robot from pointing to anything directly in front of it using only the joints present in its arms. This is a problem for tabletop interaction; although it can be mitigated to a certain extant with the use of the wrist joints, it restricts the area in which the robot can indicate objects with each hand while using only the degrees of freedom present in each of the arms. In order to overcome these limitations, a pointing pose generation scheme was implemented which used the joints in the hips and elbows of the robot to allow the robot to point to areas that it would not be able to point to otherwise.

4.2.2.2.1 Pointing Pose Generation

Due to the limitations in the degrees of freedom of the arms and the restricted range of motion, RoboThespian is unable to reproduce human motion identically. However, previous work using a humanoid robot has shown that effective gesturing can be produced with motions that are sufficiently human-like (Bremner, Pipe, Melhuish, Fraser, & Subramanian, 2011). The control scheme presented here provides a method to produce human-like motion for pointing when using humanoid robots.

A series of key poses were first identified from the preliminary HTI study. In particular, the joint positions of the shoulder and elbow were noted to vary depending on the proximity of the identified object. Hence, joint positions for the shoulders of the robot (pronation/supination and abduction/adduction) and elbows were specified for the boundaries of the required work volume (that is, the edges of the TID); these poses are shown in Figure 4.2-A and Figure 4.2-B with the white and black regions being the virtual representation of the Microsoft Surface and green arrows illustrating the direction vector from the end effector to the TID. In order to point towards a specific position inside the boundary area, interpolation between the boundary poses was used to find a pose that puts the target position within the range of locations that the robot is able to designate.

With the interpolation generating a human-like pose, it was necessary to ensure that the robot was actually pointing to the desired position on the TID. In order to accomplish this, the interpolated pose was set on a kinematic model of the robot and the required position on the TID was calculated in the coordinate frame of the robot. Hence, inverse kinematics could be used to calculate the hip and wrist joint values in order to align the robot's forefinger with the target. By maintaining the shoulder and elbow joints in the interpolated positions, the pose was kept as human-like as possible, while also reducing the problem to two degrees-of-freedom allowing for only one solution. The control scheme thus mitigates the limitations in the ranges of the joints in the arms by making use of joints in the hips to allow the robot to point to artefacts that would be inaccessible otherwise.

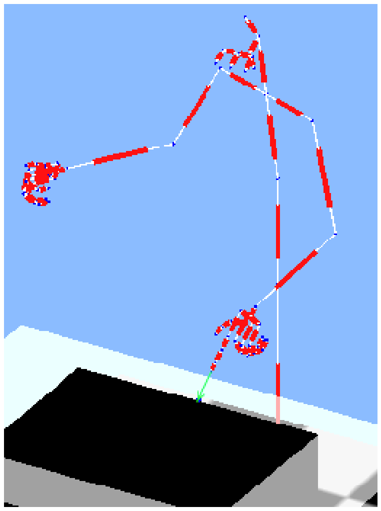

Figure 4.2-A: Pointing to a Close Point

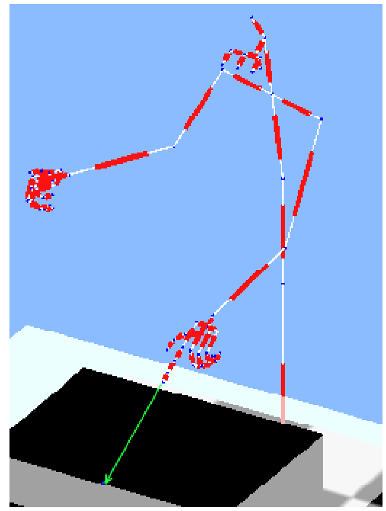

Figure 4.2-B: Pointing to a Far Point

RoboThespian Pointing (Note left arm position between figures

From the configuration obtained, the position that the robot is pointing to was calculated, using forward kinematics. The position was then checked to be within one centimetre of the target position and if it was the robot was set to that configuration. In the case where the position was found to be outside the acceptance margin, the shoulder and elbow joints were finely adjusted in order to orient the pointing to the target position. The values can be obtained using the method in section 4.2.2.5.

4.2.2.2.2 Gaze

In addition to the pointing gesture being implemented to designate a specific artefact on the TID, gaze was also implemented on the robot. Since the eyes of the robot are concentric circles of different colours being displayed on flat LCD displays, they are not very useful to provide directional gaze (Lincoln, et al., 2009). While they can be used to convey gaze, the direction information of the gaze will be limited to looking in front. This is due to the non-deformation of the view of the iris as discussed by Wang and Sung (2002). Therefore, gaze was implemented using head orientation, the other main component of gaze (Wang & Sung, 2002). Joints from the neck and torso were used in order to allow the robot to gaze at specific artefacts. To do this, a virtual vector that ran from the back of the head to the middle of the eyes was used. This vector was aligned with the target position in a manner similar to the pointing gesture described previously while the eyes remained static. The values can again be obtained using the method in section 4.2.2.5.

The gaze has been implemented in conjunction with the orientation of the finger to mirror the Human-Human Interaction (HHI), and in order to prevent adverse effects in the form of misdirected robot attention due to wrong gaze, as suggested in other studies (Emery, 2000; Goodwin, 2003; Tan, Robinson, Culbertson, & Apostolopoulos, 2010). Conversely, while the gaze directs the attention of the participant to areas of the TID, it alone does not provide enough information for the participant to identify the artefact uniquely, because of the spatial density of the artefacts.

4.2.2.3 ABB Arm (Non-humanoid Robot)

As mentioned earlier in the chapter, the non-humanoid robot used in the investigation is the ABB arm industrial manipulator. While it is not able to execute complex human gestures like the RoboThespian, it can however perform limited ‘point’ gestures such as to objects on a TID, and these gestures can be performed at high speed and with good accuracy. This is accomplished by using the powerful motors of the ABB arm, as stated in Chapter 3, to control the 6 degrees of freedom, thus allowing it to quickly assume a targeted configuration. The movements of the joints are synchronised so that all the joints of the robot start and end at the same time even if the angles that they are describing are different. This synchronisation is achieved by the robot control device by varying the maximum angular speed (using a sinusoidal velocity profile) of the different joints for each commanded movement.

In order to make the ABB arm suitable for pointing, a custom end-effector was designed and attached to it. This consisted of a thin cylindrical tube of 1 centimetre diameter and 16 centimetres in length, providing a similar capability to the pointing finger of the RoboThespian.

4.2.2.4 Trajectories of Movement

The different physical forms, ranges of motion, and joint coordination schemes result in quite noticeably different pointing gestures.

Figure 4.2-C, Figure 4.2-D, and Figure 4.2-E demonstrate how the trajectories vary. It is to be noted that in all three figures the robots are present at the exact same position around the TID. Both robots are also targeting the same point, in all three figures and the point being targeted is slightly to the left of the centre of the TID.

In Figure 4.2-C, the RoboThespian is pointing towards the target position. The pink dots represent the rest position of the main joints of the targeting arm; the indigo lines the link between these positions; the dark green lines represent the position of the arm components halfway through the motion and the yellow arrows describing the movement. It should be noted that in this case the torso of the robot is being rotated anticlockwise with respect to the base joint, hip of the robot, in order to allow the robot to not only point towards the targeted position but to do so in a human-like manner while the hand and configuration of the fingers were changed in order to produce a pointing gesture.

The use of gaze can also be seen in Figure 4.2-C, with the brown dots indicating the position of the eyes, and how it can be used in conjunction with the finger pointing.

In Figure 4.2-D, the ABB arm is being used for the same purpose as Figure 4.2-C for the RoboThespian, that is, targeting the exact same position with the pink dots, indigo lines, dark green lines and yellow arrows used for the purpose. It should be noted that since both robots have the same start and end points, the respective base joints have each been rotated anticlockwise in order to point towards the same position.

In Figure 4.2-E, an overlay of the two robots with their trajectories is shown to provide comparison with both the rest position and the direction of rotation of each robot being clearly visible.

One element, that is of particular importance, is the joint coordination between the different robots, the effect of which is represented in Figure 4.2-E by the dark green lines. The RoboThespian uses fixed velocity profiles for the joints whereas the ABB arm uses variable velocity profiles. This results in the RoboThespian having some joints already reaching their final position while the overall motion is only half way through, illustrated in Figure 4.2-E by the elbow joint almost at its target position while the wrist still having half of its motion, whereas in the case of the ABB arm all the different joints start and end as part of the motion, shown by the different joints all being at halfway through their paths of motion. This makes the motion of the RoboThespian more easily identifiable than the ABB arm since the use of general trajectory information to calculate the final end point is also used in HHI (Tang, Neustaedter, & Greenberg, 2006).

The different joint coordination of the robots caused a different perception of the motion by the participants and ultimately the pointing gestures. In the case of the RoboThespian, the motion would appear to an observer as comprising of multiple phases since some of the parts of the robot would stop moving before others, due to their joints having already reached their target positions. On the other hand, the motion from the ABB arm would appear to be happening in a single phase. This multiphase motion would make the motion from the RoboThespian not only appear more natural to the observer, closer to human gesturing (Riek, et al., 2010), but also provide a better estimation of the target position when compared to the one generated by the ABB arm. This estimation of the target position will be dependent on the number of joints in motion and the angles these joints have to move through. In the case of the RoboThespian, the number of active planes-of-motion decrease during the motion due to the different joints reaching their target position, thus suggesting a simpler estimation, by gradually confining the possible choices of zones as the motion continues. However, for the ABB arm the number of active planes-of-motion stays the same throughout the movement leading to a more complex estimation, that is, the final position can only be estimated at the very end of the motion.

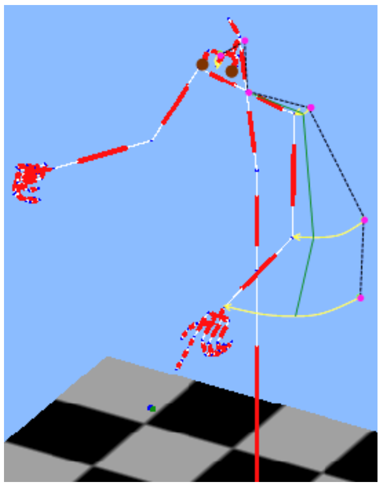

Figure 4.2-C: RoboThespian

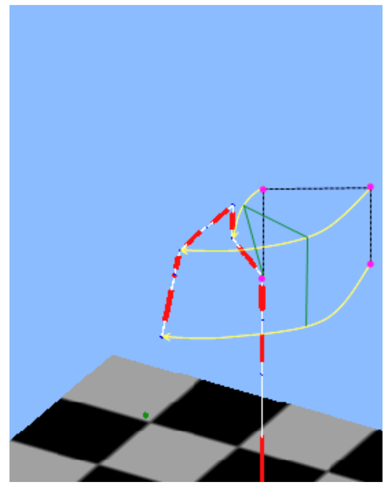

Figure 4.2-D: ABB Arm

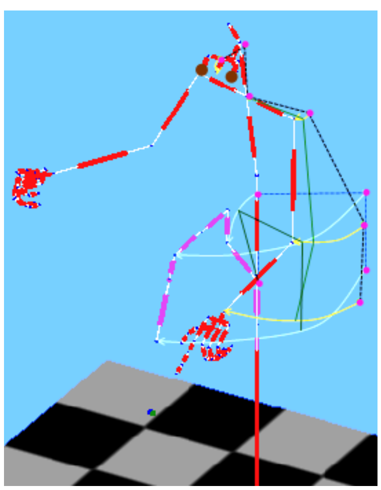

Figure 4.2-E: Overlay

Robot Trajectories

4.2.2.5 Targeting an Artefact

In this chapter, there has been several references made to the targeting system used to make the humanoid robot point, using either the arm or using the gaze, to an artefact on the TID. While there are a number of different ways to achieve pointing, the one that was used in this thesis with respect to humanoid pointing is going to be described here. Part of this description includes the logic and algorithms behind the selected method.

The method used here to target the artefact works in concert with the previously mentioned pose generation scheme. So while the scheme allows for the generated pose to be humanlike while pointing to the general vicinity of the target, this method allows for a fine tuning of joints to point to the specific target. Consequently, this method provides a value for the relative change in angle required for each joint, in order for the joint to be able to point to the target position. Using this method allows for the required change to be given while ignoring any starting position or bias that may be present on the joint already. The value received is also unsigned, meaning that the offset can either be an increase or a decrease in the value.

As with many algorithms there are a number of prerequisites for the current method to be used; this method requires the value of four important elements of the setup:

- The relative position of the target
- The axis of rotation
- The vector of alignment
- The relative position of the start of the vector of alignment from the point of rotation

Note: Contrary to touch, where the end-effector makes contact with the object, pointing can still be performed with objects that are out of reach of the entity designating the object.

Prior to applying the algorithm of the method to obtain the required change in angle it is important to simplify the problem from a 3D problem to a planar 3D problem, effectively a 2D problem across the plane. This requires for point of rotation, vector of alignment, position of the start of the vector and target position to be on the same plane. The plane in question is defined by the rotation axis, used as a normal to the plane and the point of rotation. If the values mentioned above are not lying on the plane, a simple projection can be used to obtain the values with respect to the plane. In order to obtain a planar problem as in Figure 4.2-G.

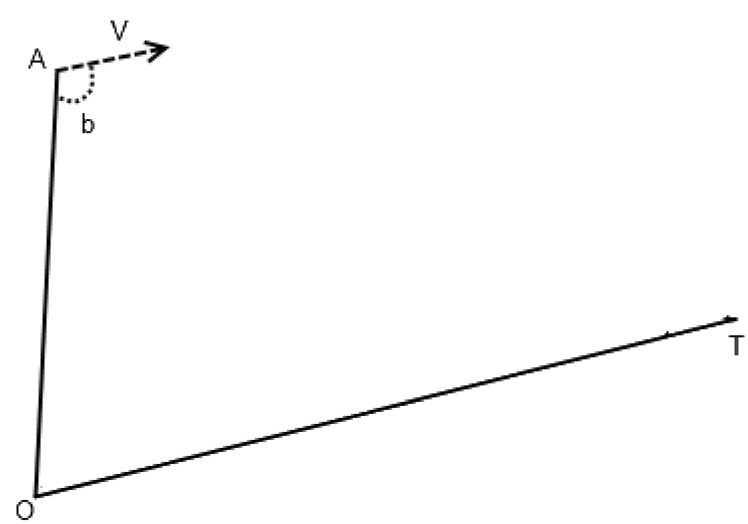

Figure 4.2-F: Targeting Problem

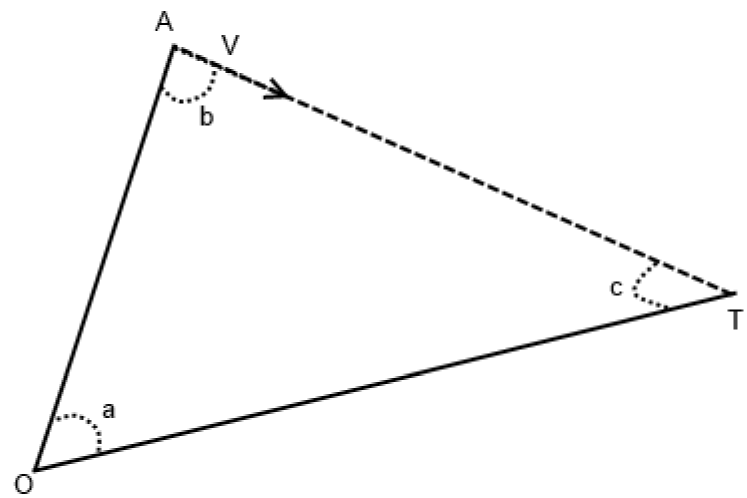

Figure 4.2-G: Pseudo-Solution Triangle

Using the four elements mentioned above, the targeting problem can be imagined as a triangle (See Figure 4.2-G) with three points O, A & T, where O is the origin around which the rotation is to take place, A is the anchor point from where the alignment starts, and T is the target position. A vector V represents the alignment vector. All these components except for point O are virtual and so when being applied to a problem a projection of each of them onto the plane of rotation may be necessary as described previously.

The triangle in the pseudo-solution can be decomposed into three main vectors: $\underset{OA}{\rightarrow}$, $\underset{AT}{\rightarrow}$ and $\underset{OT}{\rightarrow}$, where $\underset{OA}{\rightarrow} + \underset{AT}{\rightarrow} = \underset{OT}{\rightarrow}$ and V is a unit vector of $\underset{AT}{\rightarrow}$ that is, $|V| = 1$, $\left|\underset{AT}{\rightarrow}\right| = k$ and $kV = \underset{AT}{\rightarrow}$.

In order to be able to find the angle that the joint needs to be set in order for vector V to point to T, three pieces of information are required, position T with respect to O, length of side OA and angle b. Having these information, angle a and position A need to be determined. In the case where b = 90º a can be easily obtained from the following equation:

$$a = \cos^{-1} \left|\underset{OA}{\rightarrow}\right| / \left|\underset{OT}{\rightarrow}\right|$$

However, since b = 90º cannot be guaranteed, a more complex solution was required. In order to explain the solution fully, the following reference triangle will be used.

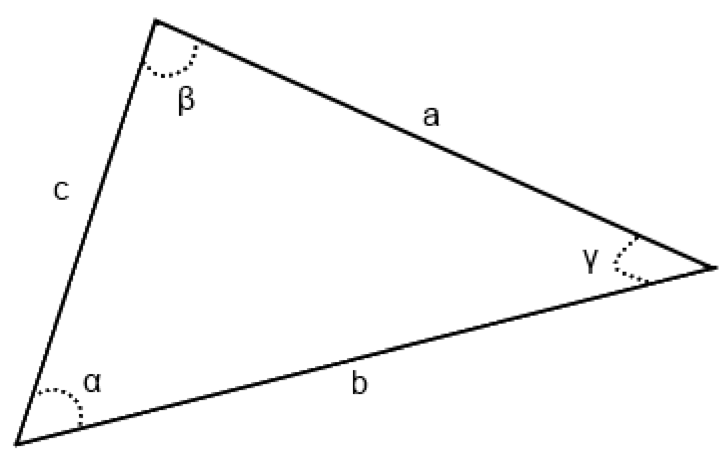

Figure 4.2-H: Reference Triangle

Together with the above figure, four equations are also necessary they are:

Equation 4.2-i: Law of Cosine

$$c^2 = a^2 + b^2 - 2ab\cos\gamma$$

Equation 4.2-ii: Law of Sine

$$\frac{a}{\sin\alpha} = \frac{b}{\sin\beta} = \frac{c}{\sin\gamma} = D$$

Equation 4.2-iii: Quadratic Formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Equation 4.2-iv: Pythagorean Trigonometric Identity

$$\cos^2\alpha + \sin^2\alpha = 1$$

Since the solution to the problem is centred on finding a value for α in Figure 4.2-G, which equates to α in Figure 4.2-H and using Equation 4.2-i: Law of Cosine above, the following equation is obtained:

$a^2 = b^2 + c^2 - 2bc\cos\alpha$ ... (1)

This can be rewritten as:

$\cos\alpha = \frac{a^2-(b^2+c^2)}{-2bc}$ ... (2)

Using Equation 4.2-ii: Law of Sine above, where:

$\frac{a}{\sin\alpha} = \frac{b}{\sin\beta}$ ... (3)

This can be rewritten as

$$a = \frac{b\ \sin\alpha}{\sin\beta} \text{ ... (4)}$$

Substituting (4) in (2):

$$\cos\alpha = \frac{(\frac{b\ \sin\alpha}{\sin\beta})^2 - (b^2+c^2)}{-2bc} \text{ ... (5)}$$

Substituting the rearranged Equation 4.2-iv in (5):

$$\cos\alpha = \frac{\frac{b^2\ (1-\cos^2\alpha)}{\sin^2\beta} - (b^2+c^2)}{-2bc} \text{ ... (6)}$$

After expanding and rearranging of the different components:

$$\frac{b^2}{2bc\sin^2\beta}\cos^2\alpha - \cos\alpha - \frac{b}{2c\sin^2\beta} + \frac{b}{2c} + \frac{c}{2b} = 0 \text{ ... (7)}$$

The above equation can be solved for $\cos\alpha$ using the quadratic formula:

Equation 4.2-v: Value of cos α

$$\boldsymbol{\cos\alpha = \frac{1 \pm \sqrt{1 - \frac{b^2}{c^2\sin^2\beta} - \frac{1}{\sin^2\beta} + \frac{b^2}{c^2\sin^4\beta}}}{\frac{b}{c\sin^2\beta}}}$$

Note: Equation 4.2-v: Value of cos α can give a maximum of 4 values for α range of $-180 < \alpha < 180$ **and is only valid when** $\left|\underset{OT}{\rightarrow}\right|$ **or** $b > 0$**,** $\left|\underset{OA}{\rightarrow}\right|$ **or** $c > 0$ **and** $\sin\beta \neq 0$.

For a step-by-step solve, see appendix section B.1

Using the appropriate value for α from the values obtained, the vector V can be aligned with the target position T. The appropriate value of α can be found out using forward kinematics for all values obtained. From this value, the required values for each of the joints can also be obtained.

4.2.3 Procedure

4.2.3.1 Main Study

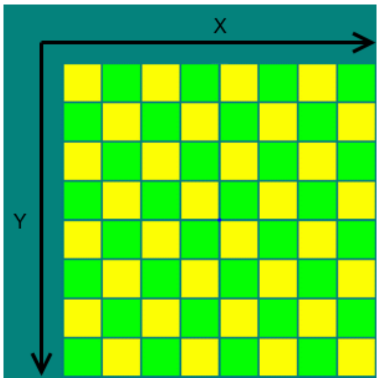

Figure 4.2-I: Grid (Arrangement)

Figure 4.2-J: Pad (Input)

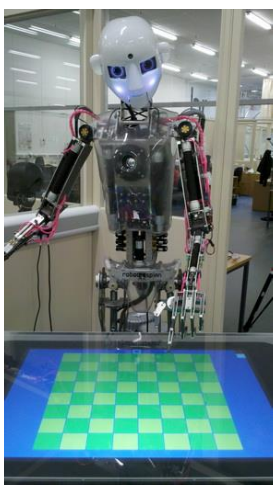

Figure 4.2-K: Tile Selection

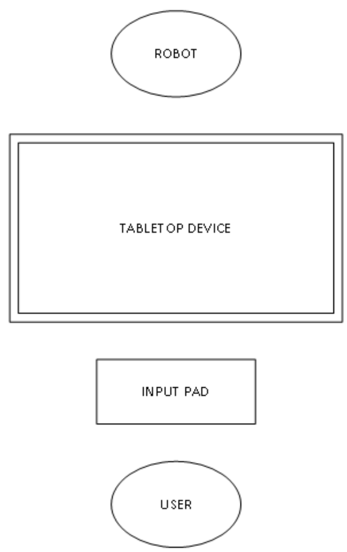

Figure 4.2-L: System Setup

In order to investigate the hypotheses stated in section 4.1.2, a set of experiments were carried out. These experiments were designed to evaluate how the use of an embodiment impacts the speed and accuracy of the selection of an object. They also investigate how the pointing gestures generated by the controller implemented for the humanoid robot fares in comparison with both the pointing from a human being and that from the industrial robot.

The robots were tasked with designating specific artefacts present on the Microsoft Surface, that is, individual tiles from a grid of alternating coloured square tiles (See Figure 4.2-I). The experiment was conducted in two distinct phases. In the first phase the focus was solely on the accuracy in identification of the tile. So, the task of the participants was to identify which part of the TID was being indicated by the embodied agent. In order to more finely assess the performance of this phase of the experiment, it was carried out using an 8x8 grid of squares each with sides of 5 cm. In order to capture the participants' responses a Launchpad[14] was used as an interface, which was configured to display specific colours mirroring the grid of tiles on the Surface (See Figure 4.2-J). Each test was started with the robot selecting a tile at random and indicating it to the participant (See Figure 4.2-K). After the robot indicated a tile, the participant was required to press the corresponding button on the Launchpad Controller to the tile they thought was being pointed to.

For the second phase of the experiment the focus was placed on the response time and how the use of the different robots affects the response time when an object is being designated. In this phase, an 8x8 grid was again used (See Figure 4.2-I) with the addition of a cursor, which was represented by a red 1 cm by 1 cm square that would appear in the middle of a tile to designate that specific tile. The appearance of the red cursor would occur about 1.5 seconds after the start

[14] Launchpad is a product developed by Ableton & Novation. ***http://www.novationmusic.com/products/midi_controllers/***

of the pointing motion in order to make sure that the cursor only appeared after the different embodiments finished the pointing motions. The cursor was used for two reasons: firstly, with the cursor present differences in accuracy between embodiments was controlled for; secondly, the cursor allows the comparison of the impact of different embodiments through comparison with a cursor alone.

The participants were tasked to press a button as soon as they saw the cursor. The response time captured in this experiment was the time between the cursor appearing on the Microsoft Surface and the participant pressing the button.

In both phases of the experiment, each embodiment was set up facing the TID, with the hand, in the case of the humanoid, and the pointer, in the case of the industrial arm, about 100 mm from the surface. This distance was selected as a safety buffer to minimise the possibility of collision and prevent any accidental damage to either the robots or the TID. Each participant was on the other side of the TID with the input pad, such that they could make a direct mapping of the tiles on the TID with the keys on the input device. Each participant took part in both phases of the experiment. As a control condition, a human was placed in the position of the robots and performed the same task. For the second phase of the experiment, an additional control condition with no embodiment was carried out, in order to determine the response time when using the cursor without an embodiment. In each phase of the experiment, the conditions were varied between subjects using partial Latin-squares.

4.3 Results

4.3.1 According to Accuracy

As mentioned previously, the experiment was carried out in two phases, with the first phase focussed on the accuracy of the pointing from the embodiments, without the use of a cursor (3 conditions) and the second on the response time using each embodiment condition together with a cursor (4 conditions). After analysis of the results obtained from the first phase of the experiment Figure 4.3-A, Figure 4.3-B, Figure 4.3-C, and Figure 4.3-D were obtained.

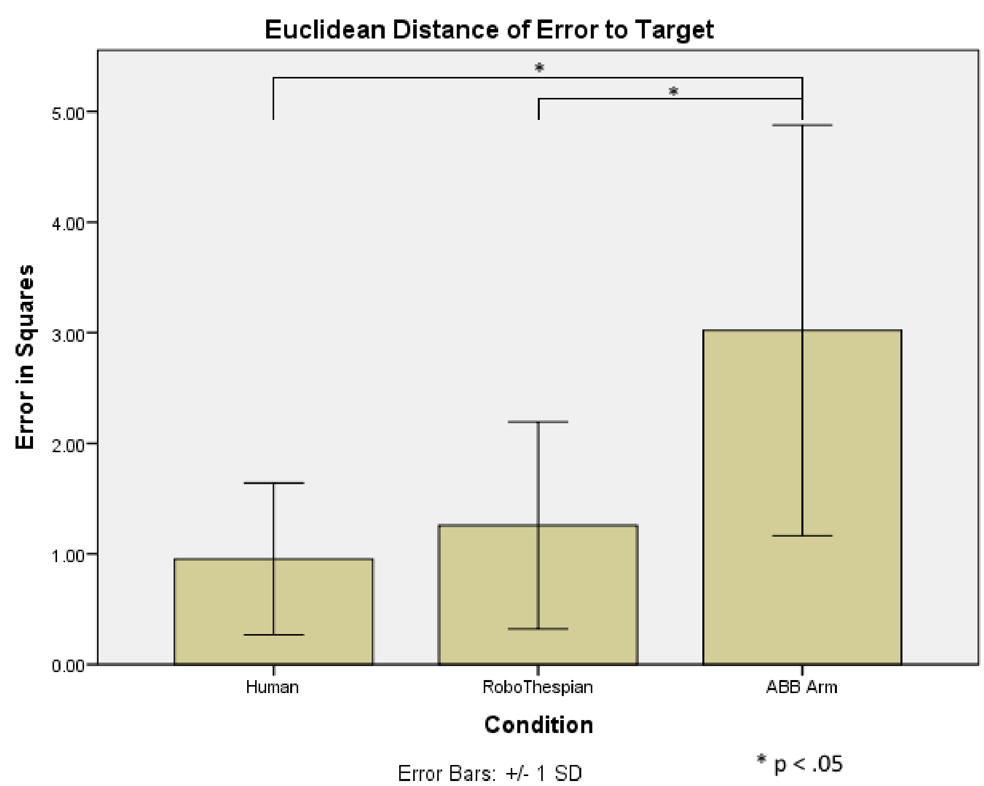

Figure 4.3-A: Euclidean Distance of Error to Target

From the data presented in Figure 4.3-A, it can be seen that the pointing provided by the RoboThespian using the described control scheme allows the participant to estimate the target tile with an error margin of 1.414 squares ($M_{error} = 1.26$), that is, a resolution of 7.07 cm. It can also be seen that in the case of the ABB arm the error is more than twice that of the RoboThespian.

Since the experiment was carried out using a Within Subject Design, that is same participants were used for each condition with the conditions being independent, and the values collected are all continuous, a repeated measures ANOVA was used to analyse the data. While the repeated measures ANOVA or rANOVA is a possible test here because of the Within Subject Design and the data collected being continuous it has a number of assumptions that need to be checked. These are the normality in the distribution of the data, and the sphericity which is the equivalent of the homogeneity of variance in the case of repeated measures.

In order to use the rANOVA to test the data, a Shapiro-Wilk's test ($p>.05$) (1965) was conducted and the result showed that the data samples were not normally distributed. Since the data normality requirement of the data was not met, a Friedman Test was conducted instead of a rANOVA. The results of the test showed a statistically significant difference, $\chi^2 (2) = 41.4, p < .001$. Post-hoc pairwise comparisons with Wilcoxon signed-rank tests were conducted, with Bonferroni correction, the results of which showed a significant difference ($Z = 5.02, p < .001$), between the use of the RoboThespian with the control scheme and the ABB arm robot, a similar difference was also obtained between the human and the ABB arm ($Z = 5.2, p < .001$) but no significant difference was found between the human and the RoboThespian ($Z = 2.29, p = .066$). The significant difference coupled with the lower value of error in the case of the RoboThespian being used as an embodiment instead of the ABB arm, proves an increase in accuracy from the implemented control scheme. Therefore, from these results it can be concluded that the RoboThespian using the control scheme provides a definite gain over the standard ABB arm industrial robot on the accuracy of the perception of pointing.

During the experiments, one observation that became quite clear was that there seemed to be a trend in the error of the tiles, with respect to their components, that is that the error in the determination of the column component was significantly lower than the row component, across the different variations of the first experiment. In order to investigate this trend, the data was further analysed with respect to each of the components of the tile location. Figure 4.3-B, Figure 4.3-C, and Figure 4.3-D represent the result of the analysis of the demand-response from the participant when the pointing was done on the 8x8 grids using the robots and the human.

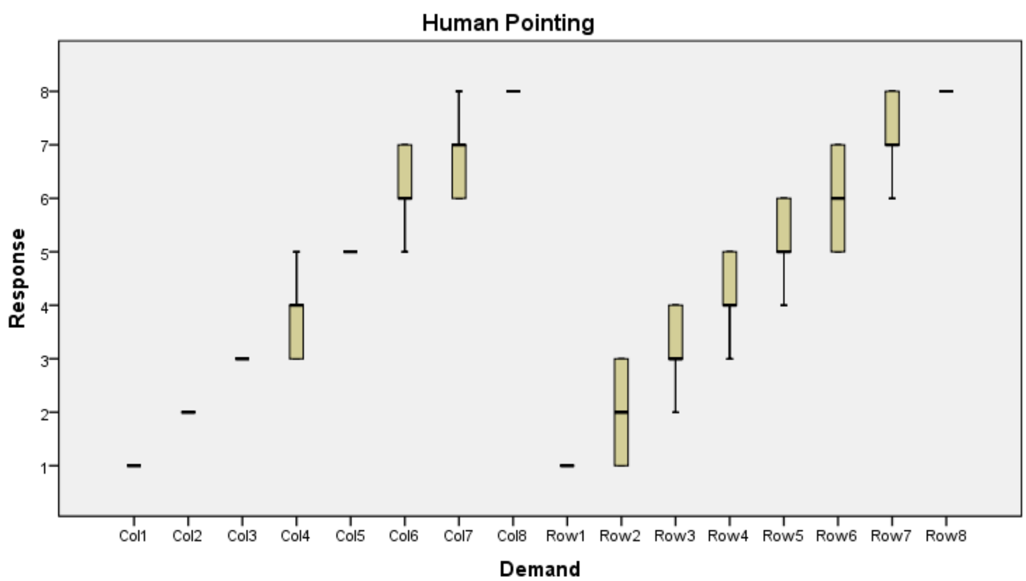

Figure 4.3-B: Human Pointing

The horizontal axes of all the graphs represent the tile indicated (indication) with respect to the components of the tile with the vertical axes representing the button pressed (response). The ***columns*** and ***rows*** represent the X and Y components respectively of a selected tile in the grid

(See Figure 4.2-I). The analysis is done with respect to the columns and rows that the tile belonged to, with the vertical lines representing the range of the input from the participant.

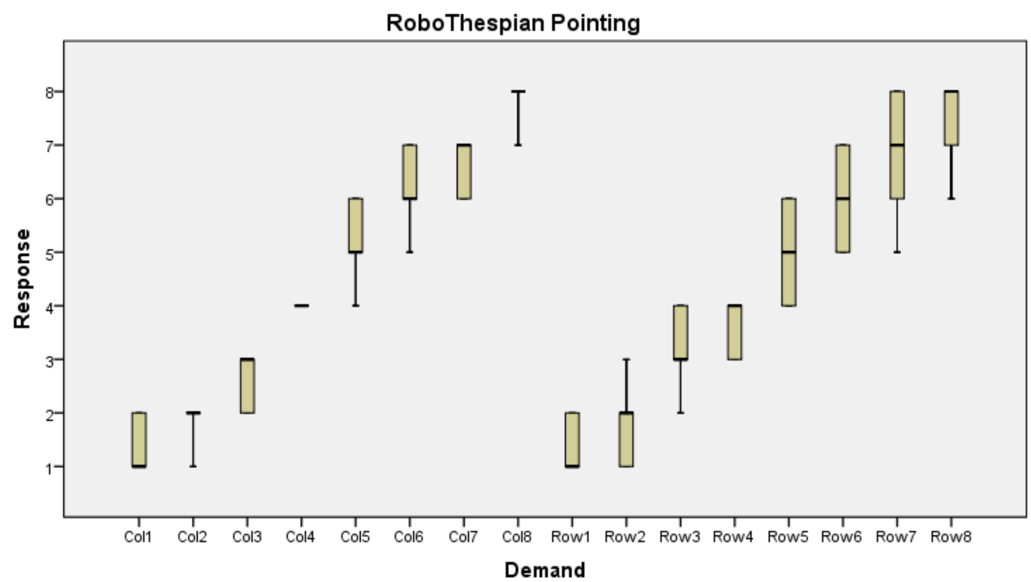

Figure 4.3-C: RoboThespian Pointing

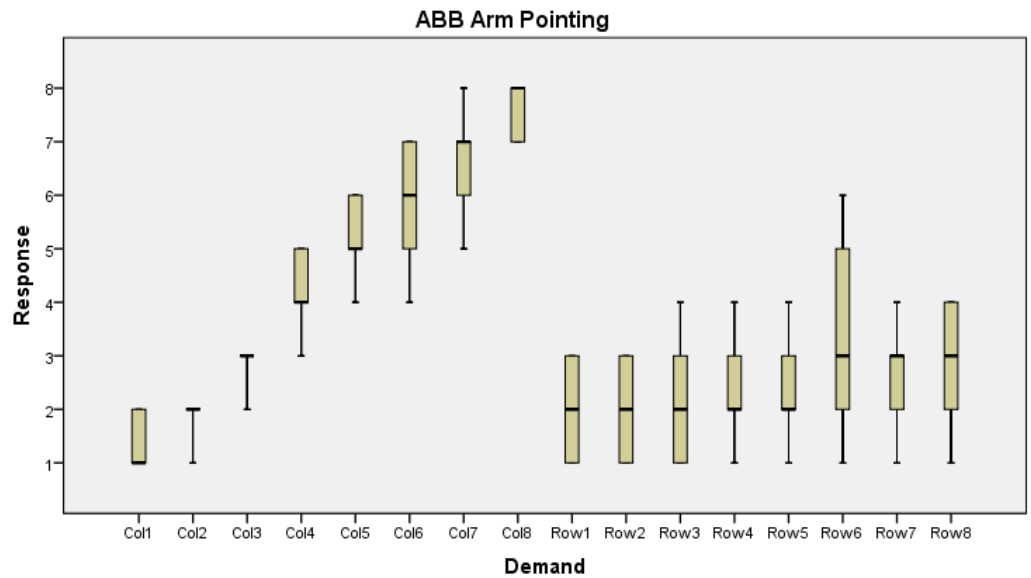

Figure 4.3-D: ABB Arm Pointing

By comparing the datasets represented in Figure 4.3-B, Figure 4.3-C, and Figure 4.3-D from the first phase, there seems to be a similar trend in the errors between the datasets obtained from the pointing performed by the RoboThespian and the one performed by the human, this trend is not as clear in the case where the pointing is performed by the ABB arm. There was however one trend that is apparent when comparing all three figures. That is, whatever the nature of the embodiment, human, RoboThespian or ABB arm, the accuracy was better when identifying which ***column*** is being designated by the embodiment when compared to the ***rows***.

A Pearson product-moment correlation coefficient was computed to assess the relationship between the error in human pointing and the one from the RoboThespian. There was a correlation between the two variables, $r = .95$, $n = 48$, $p < .001$. When the number of errors from the RoboThespian was compared to the human, it was found that in the case of the robot, the mean number of errors was increased by 1.316. There was however no proportional increase in the error zone between the two conditions, thus making the increase in error due to boundary delimitation between the tiles while the embodiment still defines the same area, this problem can be solved by the use of the cursor previously described.

4.3.2 According to Response Time

For the second phase of the experiment, four trials were again carried out for each condition of the experiment with the same 12 participants. The focus on this part of the experiment was the effect of the different embodiments on response time. The result of the experiment is illustrated in Figure 4.3-E.

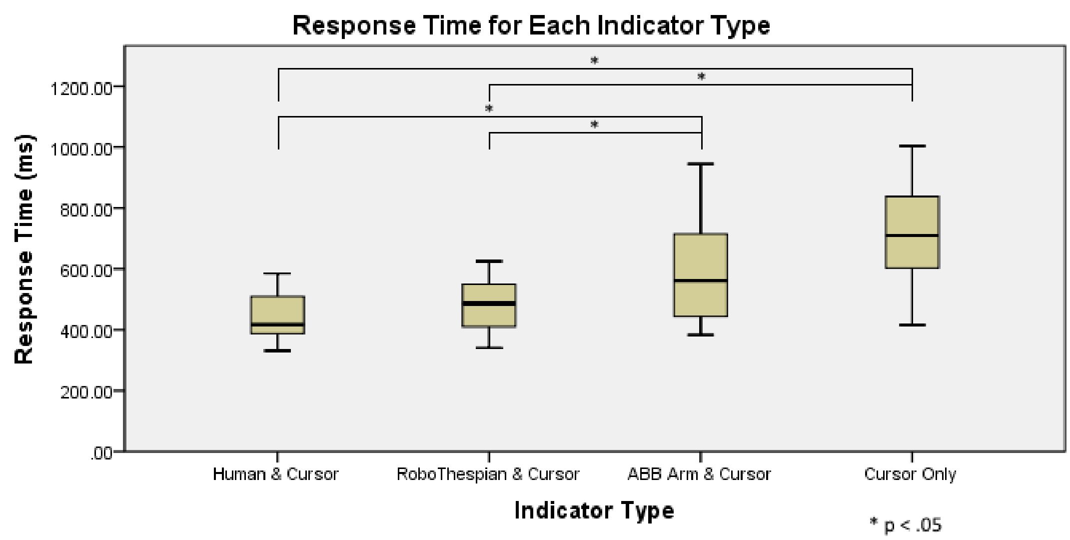

Figure 4.3-E: Response Time for each of the different setup

In Figure 4.3-E, the box and whisker diagrams represent the distribution of the response time, for finding the selected tile, when using a cursor across the different variations of the experiment. One of the observable features in the figure, even before a quantitative analysis is performed, is that the data from Human and RoboThespian variants appear to be close in terms of both range and distribution while the ABB arm has a significantly wider distribution comparable to the no embodiment variation though the median is much lower. This high variation can be attributed to high susceptibility to location, i.e., benefit was only gained for the ABB arm when the points appeared in certain locations.

Using the same rationale used in the previous section for the selection of an appropriate statistical test for the data, a rANOVA was again selected due to the data being continuous and the Within Subject Design of the experiment, if the assumption of normality was not violated. In case of violation a Friedman test would be used instead.

A rANOVA was used to analyse the data for each condition: Human & Cursor ($M = 445$, $SD = 82.6$), RoboThespian & Cursor ($M = 483$, $SD = 90.2$), ABB arm & Cursor ($M = 594$, $SD = 182$) and Cursor Only ($M = 714$, $SD = 181$).

The repeated measures ANOVA determined that mean response time differed statistically significantly between the different conditions ($F(3, 36) = 28.27$, $p < .001$). Post-hoc pairwise comparisons with dependent t-tests were conducted, with Bonferroni correction, the results of which showed that the use of an embodiment caused a decrease in response time. While there was a decrease in the response time when using the ABB arm as the embodiment, the decrease was found to be approaching significance ($p = .05$) when compared to the Cursor Only scenario. On the other hand, the use of the human and the RoboThespian as embodiments showed statistically significant results ($p < .001$) for both. Furthermore, when comparing the response time for the different embodiments, statistically significant results were obtained for both the human ($p = .015$) and the RoboThespian ($p = .048$) when compared to the ABB arm. Finally, the difference in response time between the human and RoboThespian was found to be not statistically significant ($p = .116$).

From these results, it can be deduced that with respect to the response time, the presence of an embodiment has a positive impact on the response time when the embodiment points in a humanlike manner. Moreover, on comparing the different types of embodiments performing the pointing with respect to response time a significant difference was found though the comparison between the Humanoid and Cursor conditions (Human and Cursor, and RoboThespian and Cursor) and Non-humanoid and Cursor (ABB arm and Cursor) condition. Finally, the least significant difference was found between the RoboThespian and Cursor condition and Human and Cursor condition. This indicates that when only the response time is taken into account, the human like pointing of the embodiments provides definite proof of an advantage.

4.4 Discussion

From the data obtained from the first phase of the experiment it can be seen that even without the use of the cursor the participants were able to correctly specify which column that tile belonged to and were (on average) within one square of the target row, in the case of the human and the RoboThespian. For the purposes of spatial and action awareness required in many HTI scenarios, this is likely to be sufficient for successful cooperation, showing that the RoboThespian was in this sense as successful for the task as a human.

According to the work of Oberman et al. (2007), the use of movements in robots has the potential to trigger the mirror neuron system which is used for different functions in human beings, one of them being the imitation of the action of other humans. This involves not only carrying out the action of a different person but also help to understand the action and its context without having to repeat the action. While this system has been used on other biological entities, than humans, it is not usually applied on mechanical devices, however depending on the actions being carried out by robots this may also apply to these actions (Oberman, Pineda, & Ramachandran, 2007). Its use is contingent on the robot in question behaving in a fashion that would be analogous to a biological entity (Oberman, Pineda, & Ramachandran, 2007). In this chapter, the two robots used behaved in very different ways. In the case of the humanoid robot, a human pointing approach was used, this involved the use of not only the arm doing the pointing but of the torso and the synchronisation of the gaze to mirror a person performing the action (Kita, 2003). However, the lack of head or torso prevented the same scheme to be applied to the non-humanoid robot so a different pointing scheme was selected. This different scheme made use of the whole body of the robot, making it more reminiscent to pointing from canines, which is a learned behaviour in the animal (Hewes, 1981). When taking into consideration the explanation provided above about the mirror neuron system and its use in understanding actions of biological entities in the environment, one possible reason for the difference in results between the two robot embodiment conditions can be put forward. This possible reason is that the pointing scheme used for the humanoid, which aims to duplicate the pointing of a human being, stimulates the mirror neuron system enough for it to aid with figuring out the designated artefact. On the other hand the alternate pointing scheme could be foreign enough for it not to provide the same stimulation, and be perceived as a non-biological pointing which would not stimulate the mirror neuron system as mentioned above. The impact of this could account for the difference in the values in terms of both accuracy and response time that was obtained from the use of the two schemes, each on its own robot. It is to be noted that while this can be postulated, the actual impact on the mirror neuron system would only be able to be confirmed by analysing the cerebral activity output of the participants experiencing the pointing.

When using traditional audio visual communication or long distance communication, referring to artefacts is accomplished by the use of audio to indicate the position, placement or characteristics of that object (Tang, Neustaedter, & Greenberg, 2006). In this chapter, the artefact designation is accomplished using only embodiments as pointing devices, that is without the supportive audio reference of gestural context that would usually form part of the discourse in which the pointing gesture is to be carried out (McNeill, 1985). Using these embodiments, the modality of the information, which is traditionally audio, has been transformed into a visual signal paralleling how human beings communicate. This would thus allow sensory-bridging communication if the pointing is performed to designate objects as part of a distributed interaction setting (Baranyi & Csapo, 2012). Though this is achieved with both embodiments, the designation scheme implemented in the humanoid robot more closely matches the designation from a human being and so provides for better communication.

As mentioned previously, in order to better understand the issues with identifying the correct tile being designated by the embodiments, it was instructive to decompose the error into its components, that is, in terms of rows and columns of the grid (See Figure 4.2-I). From this the main source of the error appears to be in properly identifying which row the tile belongs to (that is, the Y value of the tile), which seems to be more pronounced in the case of the ABB arm. From this, and the setup of the experiment, it can be concluded that determining a tile along the X axis of the grid is better than along the Y axis for a square grid shown by the slightly smaller inter-quartile boxes. One possible explanation that can be put forward for this trend in the data has to do with the visibility of the pointing action being carried out by the robots. In the work of Butterworth (2003), it was stated that the perceived length of the lever doing the pointing plays an important role in the identification of the artefact being indicated. In the setup of the current experiment the robots were set on the other side of the TID facing the participants, the lever of the pointing action from the robots would therefore, when dealing with the Y components, be pointing in the general direction of the participant. Consequently, from the perspective of the participant, for two perpendicular displacements, with the same magnitude but parallel to each axis, the perception of the displacement parallel to the Y axis will seem to be smaller than the one parallel to the X axis. This would be further accentuated in the case of the non-humanoid robot, compared with the humanoid one due to the higher position of the main lever on the ABB arm than the RoboThespian. The better recognition with respect to the side rather than distance from the participants has important implications for interface design, and task performance analysis. The control method implemented here has thus shown that it can be used to allow the RoboThespian to point with an accuracy level similar to a human being. Using the control method, the RoboThespian was able to portray the pointing in a manner than was significantly better that the ABB industrial robot.

According to Fitts's Law (Fitts, 1954), the average time taken to select an object increases with the distance to target while keeping all the other factors the same, a parallel can therefore be drawn to account for the above finding. In this case, the response time is a measure of the distance to target from the initial focus area. This together with the data obtained from the second phase of the experiment, showed that the presence of an embodiment, whether it being a robot or a human, has a positive impact on the reduction of the response time, this is even more evident in the case of the human and the RoboThespian. That is, the use of an embodiment complements the use of the cursor by shifting attention to the area where the cursor will appear during the gesture, decreasing the distance, and hence the time that is required to see the cursor while the cursor itself adds an improved accuracy.

The use of embodiments, or representation of them, has also been proven to add trajectory information to interaction (Fussell, Setlock, Parker, & Yang, 2003; Kuzuoka, Oyama, Yamazaki, Suzuki, & Mitsuishi, 2000; Tang, Neustaedter, & Greenberg, 2006). This is of particular importance, especially if the RoboThespian, in this case, is to eventually be used as an avatar, for a remote person to interact with the ones present around the TID. This is supported by the lack of difference between the embodiment being a human or a RoboThespian, both in terms of accuracy and response time. In the case of the presence of an embodiment, the pointing gesture from the said embodiment directs the attention of the participant to a specific portion of the screen. With the attention drawn to a specific region of the screen the distance between where the participant is focussing and the cursor is decreased. It is suggested that there is a correlation between the distance from the focus point to target position and the response time with the more pronounced error in estimating the row component in the case of the ABB arm leading to a bigger response time (See Figure 4.3-D and Figure 4.3-E).

In the case of the robot designating tiles from Rows 1 and 2 a lower response time was obtained even comparable to the values from the humanoid embodiments. On the other hand, when designating tiles from Rows 7 and 8 the values in response time obtained from ABB arm were comparable to values obtained in the cursor only scenario. This would tend to suggest that in the case of the pointing carried out by the ABB arm that while the column component was correctly estimated, the row component of the tile was not very useful except in the case of Row 1 and 2 leaving the participant having to make an effort to find out the row component of the tile while the embodiment provided an accurate information for the column component.

4.4.1 Summary

In the investigation, carried out in this chapter, it has been shown that the use of a humanoid embodiment, causes a significant improvement in indication recognition time over the cursor alone. This finding strongly supports hypothesis H 4.1 (See section 4.1.2) in the case of the pointing being carried out in a human like fashion. A similar finding was also obtained with respect to H 4.2 since a gain in accuracy was present that allowed the pointing from the humanoid robot in particular to be comparable to the co-located human condition. Moreover, from the data obtained from the experiments it can also be observed that the performance of the humanoid robot was significantly better than that of the non-humanoid robot with regard to accuracy and response time, as it is less affected by where the pointed location is on the TID. Hence, the data strongly supports hypothesis H 4.3 in terms of accuracy and also it terms of speed. Following on from this, there was no significant difference between the performance of the task for either speed or accuracy between a human or a humanoid robot with restricted range of motion using the control scheme, further supporting hypothesis H 4.3. Further, the control scheme was able to overcome the limitations of the RoboThespian indicating its usefulness in an avatar scenario. Although these results were obtained for a tele-operated intra-cognitive sensor-bridging application, the findings could also be useful for an inter-cognitive sensor-bridging application where the robot is acting autonomously.

These findings show that the control scheme allows humanoid robots, even those with restricted range of motion, to be able to indicate artefacts successfully, comparable to a human, in the context of tabletop interaction. Further, the findings imply that very close bio-mimicry is not required in a robotic avatar for tabletop interaction. Importantly, due to the inherent safety issues, and hence design compromises common in cooperative HRI (such as compliant, low power actuation schemes), it can be suggested that making those design compromises does not have a significant effect on robot utility. The ideas presented here could be extended to other humanoid robot platforms, and also have implications for robot avatar design. One such implication is that high precision and speed, and hence a high system cost of a suitable robot platform is not required, for example, RoboThespian is around a fifth of the cost of the BERTI robot platform, previously mentioned in section 3.2.1.2, that has been used in other HRI work (Bremner, Pipe, Melhuish, Fraser, & Subramanian, 2011; Riek, et al., 2010); hence, in the context of tabletop interaction, a cheaper robot can be used. This is of particular significance if such a setup is to be used in a scalable and/or commercial system, as in creating a duplex tele-presence system where each participant is embodied in each other's area using humanoid robots.

4.5 Chapter Conclusion

The purpose of this chapter was to investigate the impact of embodiment on artefact indication, for humanoid and non-humanoid entities. As part of the investigation the person observing the robot had to identify where the robot was indicating on a two dimensional interface. The results of the investigation have shown that the control scheme developed for the humanoid robot provides a significant advantage in the perception of pointing. Using this implemented control scheme, described in this chapter, the humanoid robot, not natively able to perform useful pointing, was able to mitigate its limitations and provide a pointing gesture that was on par with a human carrying out the same pointing. Having established this capability, it can now be used to portray the actions of a remote user, and its impact on group performance can be investigated.

Though the description of the findings from the preliminary study was primarily focussed on the gestures that were used as part of interaction process, pointing and touching, it was also mentioned that the members used more complex interaction, communicating verbally, while collaborating. In the next chapter, the focus will be on the type of collaboration between the different members of a group when one of the members is tele-operating a robot embodiment instead of all three members being co-located and to find out if the same difference between robots is extended to those more complex interactions.

Chapter 5 Robot Mediated Cooperation

5.1 Chapter Introduction

In the previous chapter, it was established that the pointing gestures implemented through the humanoid robot, which was inspired by human pointing gestures, was useful at portraying autonomous pointing gestures from the robot. Moreover, these pointing gestures proved to be significantly better than the normal pointing gestures that was performed by an industrial robot when the two robots were used in the same tele-pointer capacity. It is to be noted, that the pointing gestures carried out by the different robot embodiments were only performed with respect to the action that needed to be carried out. These pointing gestures only comprised of the mechanical action of pointing without the gestural context that is usually associated with the action; that is, the audio support that usually accompanies and clarifies the element being pointed to was omitted from the experience. Following from this and while taking into account that the results showed the usefulness of the implemented pointing scheme when only the action is performed, the effect of it in a full context needed to be investigated. In this chapter, the impact of the gestures being performed in their full context is going to be established by having the same robots, as in the previous chapter, carry out the pointing on behalf of a remotely located individual.

This remote pointing actions being carried out by the different robots are within the wider context of investigating the impact of the different robotic embodiments on distributed group interactions. Although the pointing gestures, being carried out by the different embodiments, are occurring in their gestural context, the gestures being portrayed by the robots in this chapter are limited to just pointing. Similar to the previous chapter two different types of robots were used to evaluate the impact of the combination of the type of robot and the associated scheme on distributed group interaction and to evaluate whether the addition of audio cues in order to support the actions has a significant impact on the trends previously observed.

5.1.1 Regarding Cooperation

When a task needs to be performed it is usually performed by a group, however the size of the group can vary and with it the structure of the group. While larger groups tend to have a rather rigid structure, where the focus is on formal roles, small groups have a tendency to be more fluid, with the focus being on informal roles (Rothwell J. D., 2001). The formal roles, as used in larger groups, are usually related with specific positions, that is individuals maintain their status within the group for the duration of the task, while informal roles are linked to functions within the group and can therefore change during the task (Rothwell J. D., 2001). Therefore, while leadership positions are assigned to specific individuals prior to a task in big groups, in smaller group, the role of the leader can be taken up by different individuals at different points in time depending on the agreement of the different members (Hare, 1994). These two group architectures affect not only the dynamics of the group but its congruity. One definition for congruity is the ability of the different members that makeup the group to cooperate during the completion of the task. However, regardless of the size or structure of the group, the inclusion of members to a group is usually to provide gains in performance, which includes not only the additional manpower, but also effective cooperative action (Liu, Laffey, & Cox, 2008). It is to be noted that, just putting the right number of people on the task does not automatically guarantee that they would be able to cooperate with one another (Campion, Medsker, & Higgs, 1993). Even if the selected people have the right expertise for the task, their lack of cooperation in performing that task may not only prove less efficient than working together but also prove disadvantageous if they hinder each other's progress (Liu, Laffey, & Cox, 2008). Consequently, since cooperation has an impact on group performance, one of the methods of gauging the level of cooperation within a group will require the assessment of the performance during a task using task metrics while keeping all other factors the same.

According to Benne and Sheats (1948), the informal roles or functions of the members of the groups can be of three types: task, maintenance and self-centred. While the first two categories tend to increase cooperation, each in its own way with the performance of the task being the primary focus of the first type and the group cohesiveness for the second; the third category tends to disrupt the group action by providing too much focus on the self at the expense of the group (Benne & Sheats, 1948). Therefore, while all three categories provide communication within the group it is only the first two that improves the cooperation between the different individuals comprising the group but also allow them to work in such a way that they help the overall effort (Benne & Sheats, 1948). They do this by actively working as part of the group while allowing the other members of the group to build upon their individual work (Liu, Laffey, & Cox, 2008). Assessing the function of each of the different members of a group is therefore an important element in the comprehension of the group dynamics.

One important role within a group is the role of leader (Hare, 1994). As mentioned previously, the role of leader differs from the position of leader in that it is not fixed to a specific member of the group throughout the completion of the task. While Hare noted the position of leader can involve different responsibilities depending on the application, the definition put forward by Benne and Sheats (1948) specifically denotes the role of the leader as helping the group to develop better and help in increasing productivity. This definition therefore encompasses both the Task and Maintenance roles as defined previously. However, the leader alone cannot fully support the cooperation and needs the congruity of the other members of the group in order for the leadership to be successful for task completion, though the success of the task itself cannot be guaranteed (Rothwell W. J., 2010). According to the congruity theory, there can be pressure and eventually lead to conflict if viewpoints differ on what the proper action to be taken is, while similar views would prevent conflict (Osgood & Tannenbaum, 1955). Therefore, one factor that can affect congruity is the size on the group, since leadership roles are assigned by the member themselves in smaller groups, the congruity will be facilitated as the leader would be someone that is agreed upon by the members. This will lead to less pressure, while in larger groups where positions are appointed, leaders may cause conflicts if the members feel that their views are being ignored (Rothwell J. , 2012). Since the size of the group plays an important role on the potential level of cooperation within a group, this leads to important design decision for the assessment of optimum collaboration within the group.

While cooperation cannot be guaranteed between members of a group, there are tools that can facilitate collaboration between these same individuals. These tools usually allow the members to be aware of what the other members are currently doing and at the same time promoting the sharing of ideas and expertise (Liu, Laffey, & Cox, 2008). These tools can range from the traditional whiteboard to new technologies that are used to promote Computer Supported Cooperative Work (CSCW). One example of such tools is the tabletop interactive devices (TIDs), which are large multi-touch screens, such as the Microsoft Surface used here (See section 3.3.1 for a description of this hardware). Even though TID technology is relatively new, it has however been the source of much attention due to its potential in CSCW setups. From the works of Tang et al. (2006) and Tuddenham and Robinson (2007), it has been shown that TIDs have a positive impact on group collaboration in setups where all the members are co-located, while the same trend for non-co-located teams are not as evident. Since TIDs have been shown to provide gains in terms of collaboration, TIDs are going to be used as part of the design of the task.

5.1.2 Hypotheses

Similar to the previous chapter, a number of sub-hypotheses based on the main ones described in Section 2.4.2 but related instead to robot mediated cooperation have been formulated. As previously mentioned, awareness is an important element of group (Gutwin & Greenberg, 1995) and in distributed group collaboration, the lack of a physical embodiment limits awareness (Tang, Neustaedter, & Greenberg, 2006). One possible way to increase awareness and thus improve group collaboration is to create a physical embodiment in the environment of the co-located members of a group, which will portray the actions of the remote user. This would potentially result in the local embodiment of the remote user increasing awareness and improving the collaboration. This results in the following sub-hypothesis:

H 5.1 The use of a robot avatar will provide significant improvement during group task performance, when compared to groups not using a physical representation.

This hypothesis also echoes with the supported hypotheses H 4.1 and H 4.2 from Chapter 4 where the use of a physical embodiment for the pointing actions, from a possible remote operator, carried out autonomously by robots were found to be significantly more useful compared to a no embodiment condition.

In the previous section, two important elements of group setup were identified as contributing to the level of interaction and affecting cooperation: functions and roles. While these elements are attributed to the members of the group, the entrusting of these to the individual members tended to be more dynamic in smaller groups (Rothwell J. D., 2001). Therefore in smaller groups, the behaviours of the group in general, and the individual role of the members, can be attributable to the actions and perceived role of each member. Consequently in the case where a robotic avatar of a remote user is used to convey the actions of the remote user, the proper conveying of the actions can impact the general role of the member within the group. Therefore, if the actions are adequately conveyed, the awareness in the local environment may be on a par to a co-located scenario and lead to the interaction being similar to when all the users are co-located, at least where cooperative behaviours are concerned. This results in the following sub-hypothesis:

H 5.2 Groups using a robot avatar will demonstrate similar cooperative behaviours to those observed where all members are co-located.

This sub-hypothesis will be further elaborated later in the chapter.

Finally, from the findings obtained in the previous chapter, it was shown that hypothesis H 4.3, was true and that the humanoid robot, using the control scheme, was significantly more useful than the non-humanoid robot at pointing. When this is applied to robot mediated collaboration, the humanoid robot is again expected to fare better than the non-humanoid robot. This results in the following sub-hypothesis:

H 5.3 The improvement during group task performance will be significantly higher in the case of a humanoid avatar, moving in a humanlike fashion, than that of a non-humanoid one.

5.2 Method

From the hypothesis formulated in the previous section, one implied requirement for running the experiment is a baseline with which the performance and interactions of the participants can be compared with. Therefore, prior to running the main experiment, a preliminary was carried out, without an embodiment but with three players co-located, to create a baseline with which the results of the main experiments could be compared. Following from this, the main study was conducted which was focussed on the change in performance and interaction between the members of the group when one of them is no longer physically co-located. This is carried out using different conditions for the representation of the remote member of the group.

5.2.1 Participants

Eighteen participants took part in the preliminary study (6 females, 12 males); aged 24 – 31 (M = 29.1, SD = 2.56). These eighteen participants were divided into six groups of three members each. Groups of 3 members were used as the small size of the group would allow for more dynamic group interactions while still providing for the different types of interactions and associate behaviours to be represented (Fay, Garrod, & Carletta, 2000).

For the main study, a total of 18 participants took part in the experiment (6 females, 12 males); aged 24 – 31 (M = 29.1, SD = 2.56). Again, those 18 participants were divided into 6 groups where each group was comprised of 3 members. That is, the two members that shared the same physical environment would provide for direct interpersonal interaction while the remote member would be able to interact with the others via remote presence interaction.

5.2.2 Material

To test the hypotheses stated in section 5.1.2, a suitable system was required. In Figure 5.2-A, a top-view of the implemented system is shown. It is to be noted that the user on the top left corner of the diagram is a remote user and is not physically present in the same environment as the remaining users. However, the actions of the remote user are being portrayed by a robot embodiment and are being used to enhance remote group collaboration.

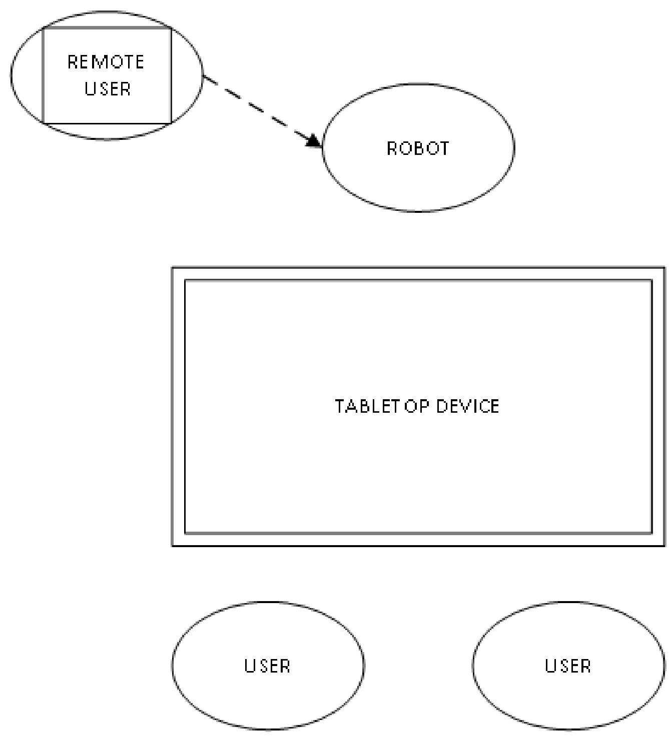

Figure 5.2-A: System Setup

Together with the above diagram, a picture portraying the participants interacting as part of the experiment is also included (See Figure 5.2-B).

Figure 5.2-B: Experiment Setting

To implement the system, a number of hardware components were required, apart from the tabletop interactive device (TID), the Microsoft Surface (See section 3.3.1), some suitable robots were also required. Both of the robots used, the humanoid RoboThespian (See section 3.2.1.1) and the non-humanoid ABB arm (See section 3.2.2.1), along with their control modules, described in section 4.2.2.2.1 were tasked with portraying the actions of the remote user. In Figure 5.2-C, the main components of the system are shown with respect to the capture and reproduction of the actions by the robots. It should be noted that the input is captured through the use of a mouse attached to a laptop, where the purpose of the laptop is to not only capture the input but provide a full interface by displaying the task. The captured input, containing the selections from the remote participant, are being translated into pointing gestures for the robots.

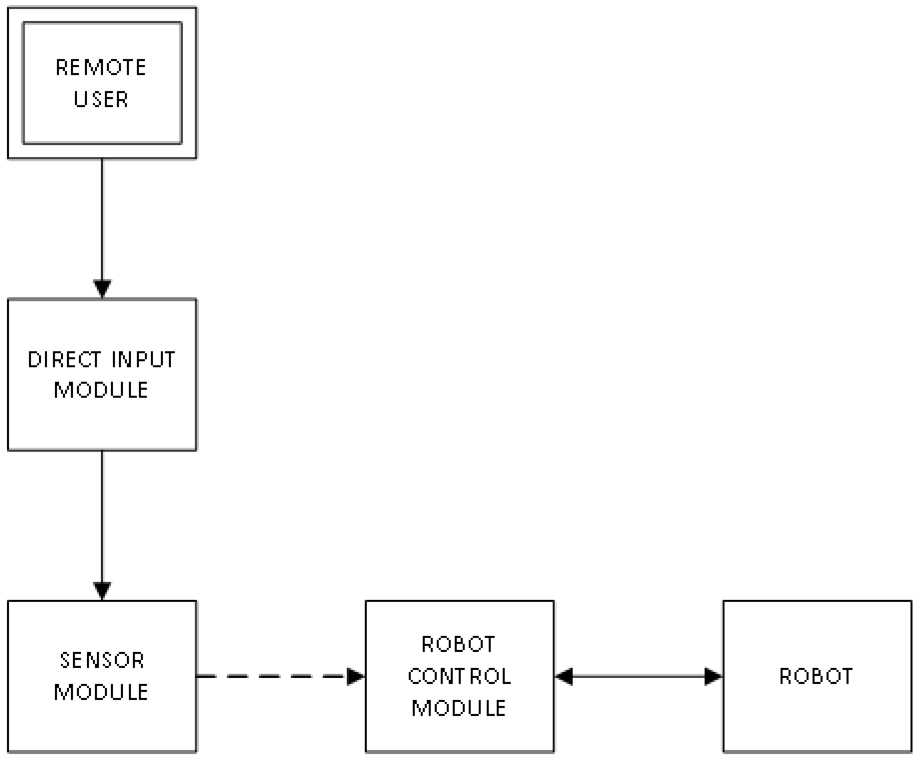

Figure 5.2-C: Capture and Output Components

5.2.2.1 Communication Peripherals

While portraying actions from remote participants is useful for proper interaction, there is still the need for the full context of the gesture which can be established by providing audio support for the interaction (Cohen, 1977). Apart from the gestural context, audio is also an important channel through which humans usually interact, and whether the members of a group are all locally present or are geographically separated, proper audio communication remains an important part of that interaction (Gaver, 1992). In order to provide audio support, all the interaction locations were fitted with microphones and loudspeakers in order to allow the audio to be captured and reproduced.

5.2.2.2 Skype

While the peripherals allow for audio to be captured and reproduced, another element was required for the audio communication to happen. This element is a software application that supports Voice over Internet Protocol (VoIP) service. The VoIP is a service that allows the delivery of voice communication over networks that use Internet Protocol (IP). The software application that was used in the experiments was Skype[15], a freely available software application that allows calls to be made between different computers or smart devices using networks that implement Internet Protocol (IP).

It is to be noted that with any communication medium, there is an inherent transmission lag to using such a software, however in the case of Skype a peer-to-peer architecture is used (Skype, 2014). Using such a distributed architecture, keeps the transmission lag to a minimum, especially in situations where the nodes are in close proximity to one another (Skype, 2014). This allows for the exchange of voice packets between the different positions even without the use of supernodes (repeaters) thus decreasing the distance that the packets have to travel to a minimum (Skype, 2014).

[15] Skype is a product of Microsoft, ***http://www.skype.com/en/***

5.2.3 Procedure

5.2.3.1 Preliminary Study

The preliminary study was carried out using the Microsoft Surface, the TID previously mentioned. For the study, six groups of three users were placed around the TID, using a similar setup as the one described in Figure 5.2-A and pictured in Figure 5.2-B but with all three users co-located and without any robots, that is with two on one side of the device and one on the other. Each group was tasked with solving a shuffle puzzle game (See section 5.2.3.2). It is to be noted that the puzzle was set in the same configuration for every group. A camera was placed next to the participants in order to capture the actual interaction between the members of the group while they were solving the puzzle. The video data obtained, from the trials, was subsequently analysed to find out the sort of behaviour that was exhibited by the different participants at specific points in time during the completion of the task.

5.2.3.2 Task

The task that was used for each run of the experiment required the participants to solve a shuffle puzzle game (See Figure 5.2-D) made up of 48 tiles set up in a 6 x 8 grid configuration. The tiles can be exchanged or swapped for one another by selecting one and moving it to another tile that is within the range, three lateral tiles, of the first one and released. Each of these swaps is considered a move. The objective of the task is to rearrange the tiles so that the original image is recreated in as few moves as possible. There is a ‘Hint’ button which shows the original image to the players and a ‘Done’ button that ends the game and evaluates the success of the operation.

This particular task has been selected since it requires the participants to point to and move artefacts on the TID. Hence, the pointing gestures previously evaluated for the chosen robot platforms are suitable for its performance. Another key feature of this task is that, while the game can be solved by only one participant, a group of participants, each responsible for specific areas or regions of the puzzle, can reduce the overall time required for completion, if they cooperate.

Additionally, the simple nature of the game, makes it easy for participants to quickly learn to play, and also facilitates performance analysis. This task is used in both the preliminary and main studies.

Figure 5.2-D: Shuffle Puzzle

5.2.3.3 Video Analysis

From an initial qualitative analysis of the recordings, it was observed that the participants interacted not only with the TID but also with the other participants. The manner in which this occurred was in drawing attention either to areas of interest or to individual tiles, using either touch, pointing or audio cues.

It was also observed that in performing the task, one of the group members often appeared to take on the role of leader within the group. The roles of different members in a group, as defined by Hare, is a status awarded by any person that fulfils the responsibilities entrusted to that status; Hare also provided a definition for the position of leader, where it was stated that the duties and responsibilities entrusted to that role can vary with respect to the application (Hare, 1994). Therefore, the term leader as it is used here specifically denotes the person in the group that is currently coordinating or instructing others as to what they need to do, or is helping them with the predefined task. Since the groups that are being used in the experiments in this chapter consist of three individuals, the overall group will tend to be dynamic with more interaction when compared to larger groups (Fay, Garrod, & Carletta, 2000). This more dynamic group structure may lead to the role of leader to not be fixed to a specific individual since the role is not being assigned prior to the start of the task. The role may therefore be taken by any member who has a best understanding of the way to accomplish the task. In Bos, Shami, Olson, Cheshin, and Nan (2004), it was stated that one of the main reasons for people to work from a different location to their work environment was that they were able to better focus on their tasks. It can be anticipated, with respect to the remote user setup (See section 5.2.2), that this will also be the case here, and hence cause the remote member to perform better than co-located group members; thus, it is suggested that the remotely located member might act as the group leader more frequently.

As a result of the preliminary analysis, a video annotation scheme has been devised to allow the hypothesis H 5.2 to be investigated. For the implementation of the video annotation scheme, a software application called ANVIL[16] was used. It is a video annotation software tool that allows video coding schemes to be implemented, used to code videos, and for the annotated videos to be compared. Due to the complexity of the interaction and the size of the group, a manual coding system was preferred to that of an automated system. Using this tool, a video annotation scheme was implemented by creating several tracks, where a track is a stream similar to the audio and video streams of the recordings and runs parallel to the other streams. Similar to the other streams, these tracks are composed of sequential packets of information each with a specific value and having fixed time coded start and end points. However contrary to the other streams, the values of the information present on the tracks need to be predefined, prior to the creation of the track. It is to be noted that the values of the packets for each track can vary along the track and that several tracks can share the same set of values.

Since the values on the tracks have to be defined prior to the creation of the track, one important prerequisite for the use of the tracks is to first define the different values of the attributes that will be used for the tracks. Therefore as a first step in annotating the videos, the videos were first viewed to identify specific actions that will be used for the annotation attributes. Furthermore taking inspiration from the work of Hung et al. among others, where dominance was identified from the use of only audio cues, a great amount of focus was placed on the audio track since it is the most available cue present in all the different runs of the experiment (Hung, et al., 2007). The annotation scheme (See Figure 5.2-E) was comprised of a track for each of the players around the table, another one for the game and a final one for the member currently in the role of the leader.

---

16 ANVIL is a software application from Michael Kipp. http://www.anvil-software.de

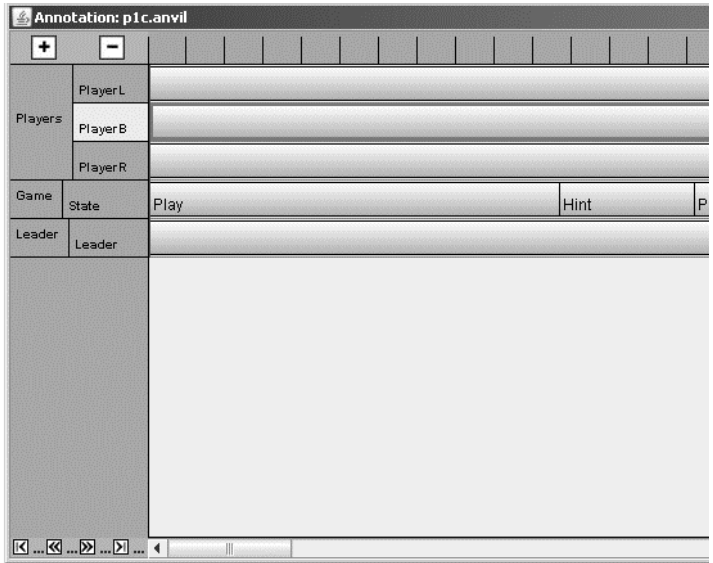

Figure 5.2-E: Video Annotation

The tracks for each of the three players share the same attributes:

- Pointing: Here it was noted what the specific player was pointing to with respect to the TID.
- Touching: Here it was noted what the specific player was touching on the TID.
- Talking: Here it was noted whether the player was talking to another player and whether they were instructing.

While pointing and touching are mutually exclusive, one use of talking was to provide the context for the gestures (McNeill, 1985) and therefore it often happened that pointing or touching was used in conjunction with talking.

The attributes for the game track are:

- State: Here it was noted what the state of the game was, whether the hint was being used or the game was running.

The attributes for the leader track are:

- Leader: Here it was noted who among the players was in the role of the leader.

- Target: Here it was noted who among the players, the leader was targeting.

- Interaction: Here it was noted what interaction type was used by the leader, whether he/she was pointing, touching or referring to features of the puzzle image.

As mentioned previously the leader role is defined by a player taking charge of coordination of the team which is characterized by the player indicating to another player where to place a tile or what to focus on. The first two attributes in this track are an adapted version of the Task Area Functional Roles as described by Pianesi, Zancanaro, Lepri, and Cappelletti, and as it pertains to the current task (Pianesi, Zancanaro, Lepri, & Cappelletti, 2007). The ‘Target’ attribute is the addressee of the conversation (Clark & Carlson, 1982) with the emphasis being drawn to by either pointing to or the speech part of the interaction.

By applying the above coding scheme to the video data from the preliminary study, similar to Takemae, Otsuka, and Mukawa (2004), a baseline was created for investigating Hypothesis H 5.2. The results of the coding scheme on the video data showed that the players relied heavily on pointing and touching for their main interaction with the TID. Referring to features of the tiles was seldom used, except when a feature was critical to the placement of the tile in a specific position relative to the rest. Both, pointing and touching, were preferred to the use of feature identification with audio support but the latter was also used in combination with either from time to time. This is particularly true for the case of the leader instructing another player on the correct placement of a tile. A similar behaviour was observed when the other players queried the leader for the correct position of specific tiles. From these results, a baseline was created and in order for hypothesis H 5.2 to hold true; similar behaviours are expected to be present in the case of the robot avatar demonstrating the action being performed by the remote user. The baseline was made up of the values of the frequency of the different observed behaviours from the annotated videos (See section 5.3.2). It is to be noted that the leader role being identified in this scheme is noticeably different from the Dominance concept of interaction in that the leader role is a dynamic attribute and can shift to different members during the task while the Dominance state would be applied to a specific individual during the interaction (Mast, 2002).

Furthermore, considering the elaborated hypothesis H 5.2, a further hypothesis can be formulated especially with respect to the role of the leader of the group where the leader is expected to be the same with respect to the baseline and in robot avatar conditions. Therefore the following hypothesis is formulated with respect to leadership:

H 5.4 Groups using robot avatars will have the same person acting as leader similar to the baseline co-located condition.

Since the same pointing gesture scheme is being used for the humanoid robot as in the previous chapter, observed behaviours from participants were used to implement the gestures being performed by the robot. These implemented gestures are used for the indication of an area of interest, while allowing the indication to be human-like.

Furthermore, metrics from the set task were identified with which hypothesis H 5.1 could be tested. The data for these metrics were captured using an automatic logging system implemented in the software. This logging system logged every action of the users together with time stamps, together with the outcome of every trial. Since these metrics are in order to test hypothesis H 5.1, they comprised of information that was captured with respect to performance of the groups in the accomplishment of the task. The metrics are therefore comprised of quantitative values for: the number of moves made by the participants, the time taken to complete the task and the number of errors at the end of the task.

5.2.3.4 Description

In order to test the proposed hypotheses, an experiment was devised whereby the same task set in the preliminary study is again performed by groups of 3 participants, but this time with one of them being remotely located from the other two; the participant who was alone on one side of the TID was moved to a remote location (See Figure 5.2-A). Since the use of the robots was to embody the remote participant and not to pretend to be autonomous robots, the participants were allowed to meet prior to the experiments being run.

The remote user (in a different room from the co-located users around the TID), was given a laptop with which he/she could interact with the game, which was simultaneously displayed at both locations. Puzzle piece moves for the remote user were captured using mouse selections. Additionally, all users were able to communicate verbally using a Voice over Internet Protocol (VoIP) system; in order to emulate the verbal communication displayed by the users when they were all co-located; this ensured that they remained able to do so.

Three experimental conditions were used to evaluate the effects of the robot avatars, one with a humanoid robot, one with a non-humanoid robot, and one without any robot, acting as an avatar for the remote user. The experiment was designed within group, that is, each group performed the same task three times. In order to compensate for the possible learning effect that performing the same task repetitively may have on the performance of the task, the order of the conditions were randomised between groups. In all conditions the tile selected and the location dragged to by users was represented by a red square on the centre of the tile (a cursor). In the robot condition, in addition to the cursor, the squares selected by the remote user were indicated using pointing gestures performed by the robots (See Section 4.2.2.2.1).

The participants were instructed on how to play the shuffle puzzle, and asked to solve the puzzle as quickly as possible, and with the least number of errors. The order of the conditions was varied between groups. The experiments were recorded, on video, while the game interactions were logged for further analysis. Three performance metrics were devised, to compare the task performance in the three conditions and evaluate hypotheses H 5.1 and H 5.3. The metrics measured were: completion time, the time required to finish the task; number of moves, the number or tile swaps required to complete the task; and number of errors at completion, the number of misplaced tiles at the end of the task.

Similar to the preliminary study, the recorded video interaction between the participants present around the TID were captured and were consequently analysed to find out if the behaviours present in the preliminary study were also present in the actual trials and so test hypothesis H 5.2.

5.3 Results

5.3.1 Task Performance Metrics

The results of the analysis of the performance data obtained from the video recordings of the interaction of the participants and the logs from the task are shown in Figure 5.3-A, Figure 5.3-B and Figure 5.3-C.

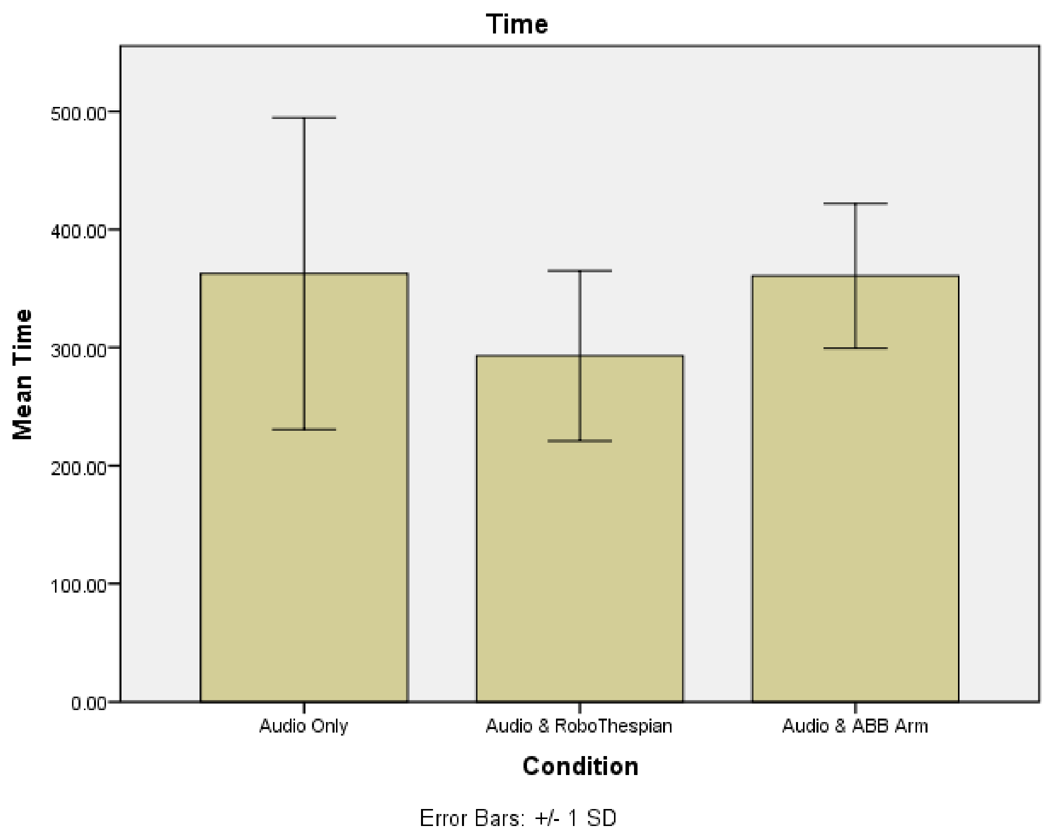

Figure 5.3-A: Total Completion Time

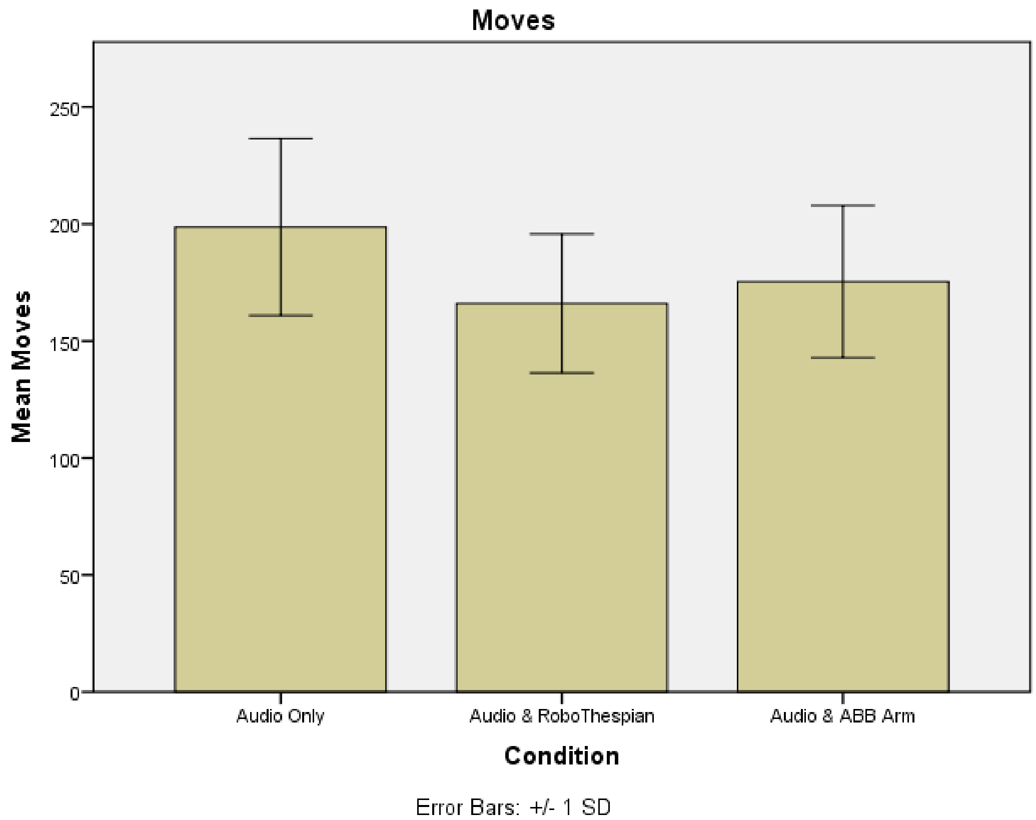

Figure 5.3-B: Moves

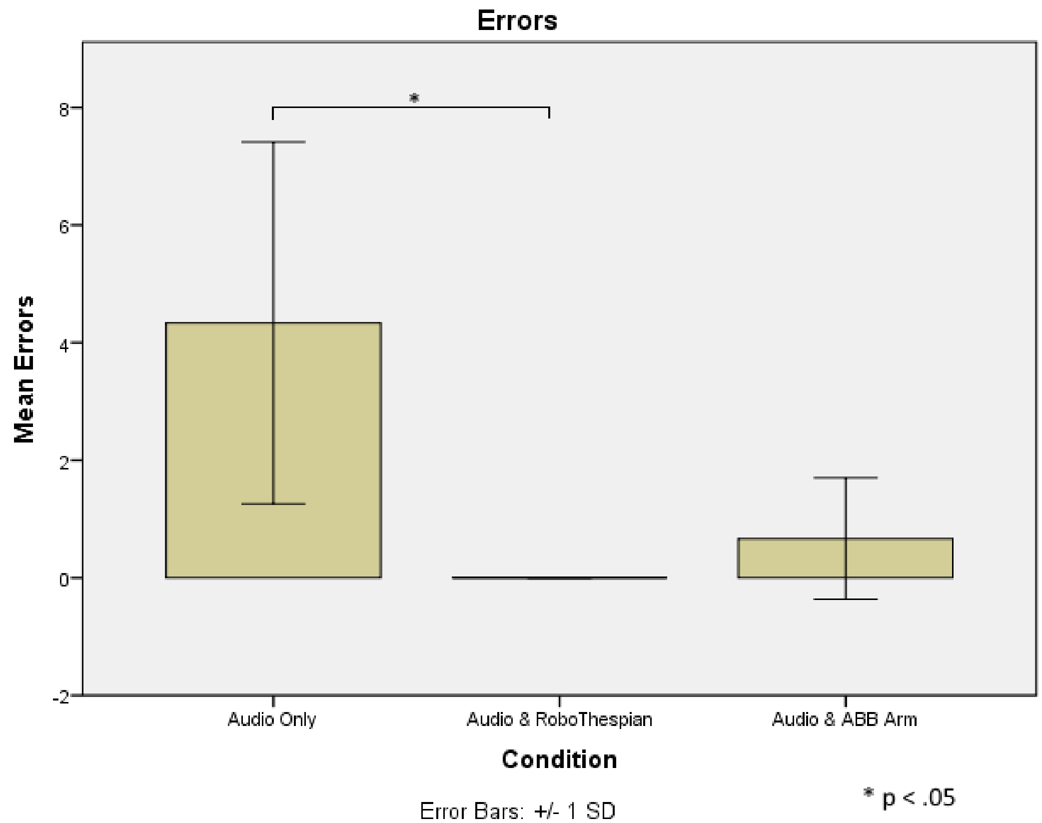

Figure 5.3-C: Errors

Figure 5.3-A, Figure 5.3-B and Figure 5.3-C show the total amount of time, number of moves (tile swaps) and the number of errors (misplaced tiles) for the group of participants for each condition.

According to hypothesis H 5.1, the use of robots as avatars would help improve the performance of the groups in the 3 metrics. So a statistical analysis was performed on each of the three metrics using a repeated measures ANOVA between each of the robots present and robot absent conditions, provided the data does not violate the required assumptions. In cases where the assumptions were violated the non-parametric Friedman test was used.

A rANOVA was used to analyse the data for each condition, with respect to time: Audio & RoboThespian ($M = 293$, $SD = 72.1$), Audio & ABB Arm ($M = 361$, $SD = 61.3$) and Audio Only ($M = 363$, $SD = 132$). The repeated measures ANOVA determined that the time taken was not statistically significant between the different conditions ($F(2, 10) = 1.25$, $p = .329$).

A rANOVA with a Greenhouse-Geisser correction, due to sphericity violation, was used to analyse the data for each condition, with respect to the number of moves: Audio & RoboThespian ($M = 166$, $SD = 29.7$), Audio & ABB Arm ($M = 175$, $SD = 32.5$) and Audio Only ($M = 199$, $SD = 37.8$). The repeated measures ANOVA determined that the number of moves was not statistically significant between the different conditions ($F(1.014, 5.071) = 2.59$, $p = .168$).

Finally, a similar procedure was also carried out with respect to the number of errors. A Friedman test was conducted and the results of which showed a significant statistical difference, $\chi^2(2) = 8.38$, $p = .015$, between the conditions. Post-hoc pairwise comparisons with Wilcoxon signed-rank tests were conducted, with Bonferroni correction, the results of which showed a significant difference ($Z = 2.06$, $p = .039$), between the use of the Audio & RoboThespian with the control scheme and the Audio Only condition. However that was not the case when comparing the Audio & ABB Arm and the Audio Only condition ($Z = 1.84$, $p = .066$).

From the results obtained it would seem that the Audio & RoboThespian condition was significantly better than the Audio Only condition only for the error metric. Although for each of the other metrics the mean of the Audio & RoboThespian condition was lower than the Audio Only condition the difference was not significant. Furthermore for the error metric, the Audio & ABB Arm condition was showed to be not significantly better than the Audio Only condition. Finally, while there was a trend for the Audio & RoboThespian condition to provide better values in general when compared to the Audio & ABB Arm condition, it did not reach significance. With respect to the aforementioned hypotheses, the values obtained partially supported hypothesis H 5.1 and H 5.3 in this part of the investigation.

It is to be noted that in the data represented in the graphs, Groups 3 & 5 differed from the overall trend. By studying the video of their behaviour, and close examination of the results, some reasons for this divergence can be suggested. In group 3, the members appeared more highly skilled (than other groups), so the robots gave them no additional benefits. Group 5 on the other hand, did poorly, high number of errors, in the audio only condition, and hence became more careful in the robot embodiment conditions, double checking a lot of the tiles.

5.3.2 Video Annotation Baseline

In order to test hypothesis H 5.2, the video coding scheme was applied to recordings of the people performing the task in the presence of each of the avatars. Also, to test the reliability of the video analysis, a second coder, not on the experimental team, coded 5 trials randomly selected from the video corpus using the coding scheme. The inter-coder reliability function, present in ANVIL, was used to calculate Cohen's Kappa, giving a K value of 0.673. This value of K shows reasonable agreement between the coding data of the different coders.

While performing the video annotation it quickly became clear that the same coding scheme devised from the videos of the preliminary study was appropriate for the current video analysis of the main study since no new behaviour was observed. While no new behaviour was observed and a similar type of overall behaviour was observed between the conditions where the robot avatars were used and when the group was co-located, there was still some noticeable changes from the baseline. From this qualitative analysis, hypothesis H 5.2 holds some truth, though there were still some differences, with the use of pointing in the Audio & RoboThespian condition being favoured, while voice description being favoured for the Audio & ABB Arm condition.

Following the brief qualitative analysis, a quantitative analysis was also performed using the video data for each condition and data obtained from the preliminary study, the baseline. A repeated measure ANOVA was again selected as the statistical test for this part of the analysis assuming the assumptions were not violated, else a Friedman test was used. For this analysis three factors were considered. These factors are: frequency of the referring to the tiles, frequency of the use of the pointing gesture, and the frequency of the remote person being the leader of the group. In this analysis the frequency of the occurrences was used instead of the count of the occurrences since the length of the sessions were not uniform.

As stated previously, the first factor that was analysed was the frequency of the referring to the tiles, this factor was selected as it was one of the two main observed behaviours from the preliminary study. The other behaviour was the designation of the tiles through pointing. A Friedman test was conducted on the data instead with the conditions being: Baseline ($M = .075$, $SD = .008$), Audio & RoboThespian ($M = .079$, $SD = .022$), Audio & ABB Arm ($M = .085$, $SD = .008$) and Audio Only ($M = .14$, $SD = .034$). The results of the test determined that, for this factor, the results were statistically significant, χ^2 (3) = 9.8, $p = .02$.

Post-hoc pairwise comparisons with Wilcoxon signed-rank tests, with Bonferroni correction, provided results in between the conditions. The results when compared to the Baseline condition, the Audio Only condition was significantly different ($Z = 2.20$, $p = .028$). On the other hand, both of the embodiment conditions proved to be not statistically significant with the Audio & RoboThespian condition being much more similar to the baseline ($Z = 0.52$, $p = .6$) while the Audio & ABB Arm condition was more different though not statistically so ($Z = 1.57$, $p = .12$). Furthermore, when compared with the Audio Only condition, the values for the Audio & RoboThespian condition proved to be statistically significant ($Z = 1.99$, $p = .046$) as did the Audio & ABB Arm condition ($Z = 2.20$, $p = .028$). Finally, no statistically significant results were obtained between the two robot conditions ($Z = 0.52$, $p = .6$).

A similar procedure was conducted for the second factor, frequency of the use of pointing, except that there were only three conditions that were compared, this is because in the Audio Only condition, there was no visible pointing gesture. The conditions are: Baseline, Audio & RoboThespian, and Audio& ABB Arm. A rANOVA was used to analyse the data for each condition: Baseline ($M = .098$, $SD = .014$), Audio & RoboThespian ($M = .283$, $SD = .063$), and Audio & ABB Arm ($M = .151$, $SD = .032$). The repeated measures ANOVA determined that the frequency was statistically significant between the different conditions (F (2, 10) = 36.8, $p < .001$).

Following from these results, post-hoc pairwise comparisons with dependent t-tests were conducted, with Bonferroni correction, the results of which showed that, with respect to the pointing to the tiles, there was a significant increase in pointing for the Audio & RoboThespian condition. Even, in the case of the ABB Arm performing the pointing action, there was an increase in the use of pointing from the embodiment, however this increase was not statistically significant, though it approached significance ($p = .055$) when compared to the baseline condition. However, in the case of the RoboThespian providing the pointing action, the increase was statistically significant when compared to the baseline ($p = .001$). Finally, a significant difference was also found between the two robot conditions ($p = .017$), with the RoboThespian providing a higher frequency of use of pointing than the ABB Arm.

For the final factor, the frequency of the remote person being the leader, a similar procedure was also carried out using four conditions, the baseline and the other three experimental conditions. A rANOVA was again used to analyse the data for each condition: Baseline ($M = .58$, $SD = .037$), Audio & RoboThespian ($M = .71$, $SD = .074$), Audio & ABB Arm ($M = .62$, $SD = .036$) and Audio Only ($M = .354$, $SD = .025$). The repeated measures ANOVA determined that the frequency was statistically significant between the different conditions ($F (3, 15) = 53.7$, $p < .001$).

Upon further analysis of the data, performed using post-hoc pairwise comparisons with dependent t-tests, with Bonferroni correction, the results showed that, with respect to the remote participant being the leader of the group, having an embodiment accounts for a greater frequency of the remote participant being the leader. From a quick glance at the above means for the different conditions, it can be seen that for the no embodiment condition, the mean is much lower. After the test, it was found that this difference in mean was statistically significant with each of the other embodiment conditions, RoboThespian ($p = .001$) and ABB Arm ($p < .001$), including the baseline ($p = .001$). While, this difference with no embodiment was significant, there was no significant differences between the embodied conditions or the baseline. So while the RoboThespian embodiment condition had a higher frequency, that value was not statistically significant from the baseline ($p = .106$). A significant difference was also absent between the Audio & ABB Arm condition and the baseline ($p = 1$). Finally, when comparing the values for the two robot conditions it was found that there was an absence of statistically significant difference between the two embodiments ($p = .324$).

In this section, the data obtained from the annotated video recordings was analysed. After testing for reliability between the coders, three factors were identified and consequently the data was processed in terms of frequencies. After statistically analysing each of these three factors, with respect to the conditions of embodiment and with the baseline, it was found that for the first factor, referring to tiles, the absence of an embodiment caused a sharp increase in the use of the referring behaviour. At the same time, the data obtained for the Audio & RoboThespian condition was much closer to the baseline condition than the Audio & ABB Arm condition, although none of the conditions proved significantly different from one another.

5.4 Discussion

In this experiment, two different result sets were obtained, one linked to the metrics captured using the logging system of the task and the other from the annotation of the video data captured of the participants performing the task. Following from the findings of the previous chapter one expected trend is that the actions depicted by the Humanoid robot will have a positive effect on the overall cooperation between the different individuals that are part of the distributed group. This trend can be confirmed from the data obtained, as it can be seen that the use of the Humanoid robot as an avatar for a remote member is beneficial to the successful completion of the task. While this is hinted at in the video analysis, it is particularly apparent in the case of the errors, the lower number of errors being achieved by the groups when the humanoid robot was used indicate that the groups are either more cohesive or focussed. Similarly to the previous chapter, the use of the non-humanoid to depict the actions, while not as useful also provided some advantages over the non-representational condition, and provided some benefits to the performance. Finally, as to be expected the inclusion of the full context for the gestures the differences in between the different conditions have been reduced since the audio compensates for the lack of information present in the action part of the gesture.

One of the findings that is related to the distributed nature of the group is the increase in performance when one member of the group was remotely situated. On comparing the lowest performance condition, the audio only condition, with the preliminary co-located group condition, it was found that there was a slight increase in performance for the error performance metric. This increase is much more noticeable in the embodied conditions with the highest being obtained when the pointing actions are being represented via the humanlike gestures of the humanoid embodiment. One possible explanation for this trend is from the works of Bos et al. (2004), where it was mentioned that one of the notable advantages of working away from the work place was the ability of the person to be more focused on the task at hand. While this gain from being remotely situated was not massive in the performance of the task, there are further factors that could account for it. In the audio only condition, there is a factor that is having a negative effect on the performance, that is there is no physical representation of the remote user for the other co-located group members (Bos, Shami, Olson, Cheshin, & Nan, 2004; Tang, Neustaedter, & Greenberg, 2006). This issue is closely linked with workspace awareness (Gutwin & Greenberg, 1999), since the lack of a physical presence can affect the awareness of the co-located participant about the remote participant. This can lead to the remote members being either ignored or excluded from the group collaboration (Bos, Shami, Olson, Cheshin, & Nan, 2004). While this is not directly shown in the results of the experiment, this may be a contributing factor to the significant difference between the “no embodiment” and the embodied conditions of the remote person being the leader of the group.

It has been suggested that the impact of an avatar is dependent on the level of interaction being displayed by the avatar for the task at hand. For the current task it seems that the designation of the tiles being manipulated and the way they are being handled, makes the work visible to the co-located users (Liu, Laffey, & Cox, 2008), and thus provide enough awareness to allow the remote user to be effectively part of the group. This is further supported by the presence of statistically significant differences between the humanoid robot and the no robot conditions for the number of moves and the number of errors metrics, for the time for completion though the difference was only marginally significant. This result can also be due to the fact that the use of embodiments to portray remote members of distributed group have been shown to have an effect on social interaction (Kose-Bagci, Ferrari, Dautenhahn, Syrdal, & Nehaniv, 2009) and with the positive impact of having humanlike motion from a humanoid robot acting as a contributing factor (Hossen Mamode, Bremner, Pipe, & Carse, 2012).

The positive impact of having a robot avatar portraying the actions of a remote participant was also present in the analysis of the coded video data, where it was found that having an avatar, irrespective of its form, allowed the group to engage in a manner similar to when all the participants are co-located. This involved the use of similar co-op behaviours and in the same proportions. However, it should be noted, with respect to the use of pointing gestures, that the frequency for the Audio & RoboThespian condition was significantly higher than the other conditions.

Therefore, with respect to hypothesis H 5.2, the results obtained for the three factors, tend to prove the hypothesis, whereby the values obtained from the robotically embodied remote operator showing no significant difference from the baseline for the first and last factors. For the second factor, the support is less present. For this factor, the frequency of pointing for the humanoid robot condition was significantly higher than in the baseline condition though that of the non-humanoid robot was not significant. So for this factor, hypothesis H 5.2 is only partially supported.

After hypothesis H 5.2 was further elaborated a following hypothesis, H 5.4, was formulated according to which the use of robotic embodiment would allow the same person to be the leader as when they are all co-located. From the data obtained from the video analysis, it would seem that hypothesis H 5.4 is supported as the difference in between the robotic embodiment conditions and the baseline was not found to be statistically significant. Furthermore, the difference between the no embodiment condition, Audio Only, and the other, embodied, conditions were found to be significant.

5.4.1 Summary

In this part of the investigation, it was shown that the use of a robotic embodiment for the representation of the actions of a remote member can have significant benefits for group interaction. This positive effect of the use of robotic embodiments was even more pronounced when the pointing gestures were being produced by the control scheme that worked in conjunction with the humanoid robot to produce humanlike pointing gestures. This result parallels the findings from the previous chapter where the humanlike gestures generated for the humanoid robot proved significantly better than the pointing gestures from the non-humanoid which also showed the gains from using an embodiment. Similar to the previous chapter, even when the humanoid robot was used in a limited capacity, with the robot just used to portray the actions being taken by the remote user, the gain in terms of accuracy and the contribution to group collaboration is enough for it to have a statistically significant impact on the interaction. The results obtained from the experiments carried out suggest the veracity of hypotheses H 5.1 and H 5.3.

Analysis of the video data, using ANVIL annotation software, allows for the different conditions to be compared and contrasted between each other and the baseline. From the comparison with the baseline it would seem that similar behaviours were observed. While they were largely similar, contrasting between them showed some noticeable differences. In fact while no new behaviours were witnessed in the three tested conditions compared to the baseline, a specific kind of interaction was favoured for each condition.

In the case where no physical representation of the remote actions were used, a sharp increase in the use of audio was noted. While the absence of a physical representation would partially account for this observation due to the members relying mainly on audio to coordinate their actions. The decrease in the associated level of performance with the amount of time that the remote person served as the leader, which was much lower in that case, could indicate a decrease in the overall group congruity.

On the other hand, in the case where the pointing was locally represented using the humanoid robot and its control system, no noticeable change in the use of audio from the baseline was found, however an increase in the amount of direct interaction with the tiles by the remote user was noted. Moreover, the participation of the remote user was increased in that particular case, with the amount of time the remote person spent as the leader was noticed to be the highest in this condition. It can therefore be surmised that the hypothesis H 5.2 holds to a certain extent while showing some marked differences in between the different representation conditions.

With respect to the group organisation and dynamics it was found that hypothesis H 5.4 was also supported by the results, obtained from the analysis of the video data of the experiments. The difference from the co-located baseline condition was not found to be significant when compared with the robotic embodiment conditions. On the other hand the non-embodied condition was found to be different from any other condition as far as leadership is concerned.

This part of the investigation highlights the utility of a robot embodiment, especially one that portrays and mirrors the localised actions of a remote participant, in the context of cooperative interaction. The positive results obtained from the use of such embodiments, on the nature of both the task and the group indicates the merits of further investigation of these ideas. Though the task used in this chapter is relatively simple, it nevertheless fosters collaboration, with congruity within the group having a positive effect on the overall results. Moreover, the use of small groups of people, with dynamic roles of the different members allows for the embodied participant to become a more active member in the group dynamics.

5.5 Chapter Conclusion

The focus of this chapter was on the impact on collaboration within a small group of people when the actions, of a remote member of the group, are portrayed using different modalities. These modalities were in three forms, using only audio with no embodiment, using a humanoid robot where the pointing is being generated to mirror human pointing and finally using a non-humanoid robot using an unaltered pointing scheme. As part of the investigation, the task that the participants had to perform was to solve a simple shuffle puzzle while the actions of the remote participant were represented to the co-located group of participants using the different setups. The results of the investigation have shown that the presence of a robot for the portrayal of the actions has a significant impact on task performance. This is even more apparent in the case where the humanoid robot portrayed the pointing actions using the specific control scheme. The results seem to indicate a net advantage with respect to that form of representation. Even when using a limited gesture base, which was further reduced to a specific type of pointing per embodiment, as used in this chapter, it was found that the robot embodiments provided a higher level of collaboration than when there was no physical representation. Since the types of gestures used were severely restricted in this chapter as a next logical step in the investigation, it is important to expand the gesture base of the robots and to find out what kind of gestures can be produced by the embodiments. This involves not only identifying the possible types of gestures but also formulating a scheme that will allow the gestures to be produced. The following chapter is therefore focussed on the investigation of such gestures when they are being captured directly from the remote operator, that is, without requiring active input.

Chapter 6 Robot Gesturing

6.1 Chapter Introduction

In the previous chapters, gesturing was implemented in different types of robots, however the type of gesturing that was implemented was only one type of gesture, pointing gesture which falls under, the deictic group of gestures. The result from the previous chapters showed that the use of this type of gesture in conjunction with the embodiments were quite useful, with the one that was implemented for the humanoid robot in order to allow it to point in a humanlike manner, even more so. Even the normal pointing gesture from the non-humanoid robot proved quite useful when compared to a situation without any embodiment. While in both chapters, the different robots were both used to portray the pointing gestures, their purpose were quite different. In the first case, the pointing gestures from the different robots were used as part of an autonomous system and without, the accompanying context; in the second case, the gestures were used within the gestural context and supported with audio. The result of the use of the different gestures showed that the finger pointing gesture generation scheme that was implemented for the humanoid robot was successful in portraying human pointing gestures, and also that the gestures from this scheme was to varying degrees better than the normal pointing gestures from an industrial robot. However, as pointed out in Section 2.2.4, deictic gestures is just one class of gestures. In this chapter, a humanoid robot is going to be used to implement and evaluate other types of gestures that are used in interpersonal communication. The gestures that are used are going to be implemented without any accompanying audio to test how successful the participants are at finding out what gestures are being portrayed. It is to be noted that contrary to the previous chapters where a scheme is used to generate the gestures autonomously, the gestures in this chapter are captured from an individual and modified in order to allow the robot to be able to portray them in a useful way.

6.1.1 Regarding Gesturing

One important element of human communication is the use of gestures to supplement the information conveyed in speech (McNeill, 1985). According to Kendon, both language and gestures are one process through which human communication is enabled (Kendon, 1980). However, the ratio of gesture to speech can vary quite a bit as well as the purpose of gesture in the communication which is illustrated by the continuum of gestures proposed by Kendon (McNeill, 1992). This continuum is one of the different ways in which gestures, which in this case refers specifically to the use of hands and arm to convey information, can be classified (McNeill, 1992). According to the continuum, in certain contexts, gestures become even more important as they are used to completely replace other forms of communication, such as sign language (McNeill, 1992). While there are other examples of gestures completely replacing speech such as semaphore, the focus here is on co-located interpersonal communication and so in this chapter the gestures that are going to be examined are those that are used in such a situation.

While gestures play an important part of co-located interpersonal communication, the use of gestures when the participants in the activity are not co-located has been observed to decrease (Cohen, 1977). According to Cohen and Harrison (1973), the gesture that accompany conversation which Ekman and Friesen (1981) refer to as Illustrators are less used in cases where the participants understand that the others cannot see them. As stated in section 2.2.4, one important element of gesturing is the visibility of the gesture to the other participants. This visibility of gestures is highly dependent on the medium of communication, that is, using certain communication media such as audio only conversation over a telephone decreases considerably the use of gestures (Cohen, 1977). Since gestures often accompany certain elements of speech, having only speech requires other elements to be used to convey all needed information or risk the other party getting confused (Ito, Hayakawa, Hotokata, & Terada, 2003). One important element to take into account in the above explanation is that the co-location of the participants in an environment is not limited to the physical environment, that is, gestures have been used to assist communication where the participants were physically distributed but virtually co-located (Hindmarsh, Fraser, Heath, Benford, & Greenhalgh, 2000).

Consequently the use of gestures as tools of communication has been expanded from situations where they were only used in interpersonal communication to being an interactive medium in communication between humans and robots (Calinon, D'halluin, Sauser, Caldwell, & Billard, 2010; Lemme, Freire, Barreto, & Steil, 2013). One of the most universal type of gestures implemented in robots is deictic gestures, these gestures have been implemented in both humanoid (Lemme, Freire, Barreto, & Steil, 2013) and non-humanoid robots (Kuzuoka, Oyama, Yamazaki, Suzuki, & Mitsuishi, 2000), with varying degrees of success. However, recent work has also been conducted in order to enable robots to understand some gestures from a human being and also to produce their own gestural response back (Calinon, D'halluin, Sauser, Caldwell, & Billard, 2010).

As stated previously, the visibility of gestures is important but that visibility is not only in terms of form but also in term of context, that is the gesture has to be visible within the right context (Cassell, 2000). Therefore one important point to take into account when using gestures in conjunction with speech is the synchronisation of the, gesture and audio, streams. Either when used between human beings or when involving other agents the synchronisation of gesture and speech is crucial for proper communication and a loss in this synchronisation can lead to confusion and a breakdown in communication due to lack of or wrong context (Cassell, 2000). When this is applied in the case of a robot performing the gesturing, the robot has certain requirements. These requirements ideally involves being able to move through the same ranges for each of the degree of freedoms and also being able to match the speed of execution of the gesture. This is in order to accurately reproduce or even mimic the original gesture. The mimicking of gestures captured from human beings by robots has involved both humanoid (Calinon, D'halluin, Sauser, Caldwell, & Billard, 2010) and non-humanoid robots (Neto, Pires, & Moreira, 2010) and the purpose of which range from social Human Robot Interaction to controlling industrial arm type robots to entertainment (Hasanuzzaman, et al., 2004; Neto, Pires, & Moreira, 2010).

One important element of that type of implementation of gesturing is the scaling factor of the gesture to make it appropriate for the robot. For example, if a gesture is copied from a 6 foot tall human to a 2 foot high humanoid robot with human proportions the range of the gesture will be severely affected. That is, a similar gesture is reproduced according to the body pose but with the net displacement being drastically affected. This scaling factor is not much of an issue when dealing with an adult human sized robot, however allowing such a robot to move at high speed raises some serious safety issues which may prevent the robot from being operated when in close proximity to human beings (Haddadin, Albu-Schäffer, & Hirzinger, 2009; ISO, 2006).

6.1.2 Implementing Robot Gesturing

The approach to mapping gestures selected is very practical as it uses an already available robot as embodiment and modifies the gestures to fit it. However as stated previously the use of normal gestures on a human sized robot raises some concerns due to the travel speed of the robot and the impact energy in any possible collisions that may ensue. While a suitable robot, for safe Human Robot Interaction, was used an implemented gesture modification scheme that would allow gestures that have been captured to be mimicked was implemented. As mentioned previously, one important element to bear in mind is the synchronisation between gestures and the accompanying audio stream. According to Cassell, when dealing with virtual embodiments, a less than 200 ms gap between the avatar starting the gesture and the corresponding word being uttered was found favourable (Cassell, 2000); hence as part of the investigation special attention was given to the implementation of an approach where such synchronisations is possible for that type of humanoid robot.

The proposed scheme that was investigated made use of scaled reproductions of the motions captured from an operator as a solution to the speed difference between the motion of the operator and the robot avatar. One of the advantages of this scheme is that it allows for a greater freedom of expression in terms of the gesture from the operator. This is in contrast to other schemes that make use of gesture classifiers, that is, first analysing the gesture that the operator is attempting to produce and then triggering the production of associated pre-programmed gestures. This approach severely limits the range of gestures that the robot is able to produce and since the gestures have to be analysed first there is an added disadvantage of increasing the latency time, i.e., a delay between operator gestures and their reproduction.

Even with the proposed scheme, one key feature to remember is that compared to virtual embodiments or overlays where the production of the movement or gesture is instantaneous, there will be an inherent delay between capture and execution. That is, in the case of virtual embodiments the representation of the movement can be at the target position immediately when the command is sent whereas using robots present in the real world that is not possible. A robot has to move a particular section of its body from a rest state to the specified target location and that travel time, combined with the slower movement speed of most robot platforms creates added delays to the production of the action.

One possible approach suggested by Sirkin and Ju (2012), is to have a more moderate motion on the robot from the one that is captured. Moreover, Sirkin et al. (2012) made the distinction in the capture of gesture data from an explicit capture device, for example the manual rig used in Adalgeirsson and Breazeal (2010), to a more implicit one and also offering advantages as to why the implicit data capture method would be more beneficial for interaction (Adalgeirsson & Breazeal, 2010). These two ideas, the more moderate motion proposed by Sirkin et al. (2012) and the implicit data capture method of Adalgeirsson and Breazeal (2010), have been used here. The system described here would be tasked with capturing data implicitly from a person and then reproduce the modified gestures on an embodiment. The advantages of such a setup include the capture a wider range of motion and a more natural motion. At the same time, the smaller range of motion would allow the context of the information, gesture and speech, to be preserved by having the moderate gestures be in line with the words.

Another important factor to consider when using robots is the transmission delay. These delays have a severe impact in the context of tele-operation. The approach usually used in tele-manipulation of objects, the “stop-and-wait” strategy, that is, waiting for the movement to catch up with the demand, is ill-suited to the context of speech accompanying gestures (Bohren, et al., 2013). According to Torok, Asboth, Honbolygo, and Csepe (2012), a delay of 200 ms is favoured when a person is dealing with both visual and audio inputs and the above strategy will affect the interactivity. It is therefore important to make sure that the delay in transmission is minimised in order to preserve the context of the communication, allowing gestures to occur in a timely manner in order to be able to properly support the accompanying gestures and minimise confusion.

Instead of using the “stop-and-wait” strategy, previously mentioned, a different strategy that would allow the gestures to fit inside the time frame of the speech would be more adequate for the transmission of gestures accompanying speech. Since the speech cannot be delayed without causing severe disruptions in the interaction, the only possible way in which both speech and gestures can happen within the same time frame is for the gestures to be shorter in terms of time. The physical limitations in terms of movement only allows for the range of motion of the gesture to be decreased in order for the gestures to fit within the time frame.

6.1.3 Gesture Selection

As mentioned in the previous section, gestures form an important part of human interaction (Kendon, 2004). The range of gestures that a human is able to perform is quite large (Kendon, 2004) and some gestures require the presence of a number of body parts working together to be performed optimally (Goodwin, 2003). Taking these points into account, it would be difficult to find a robot that is able to perform the full range of human gestures. However, the concept of using robots to perform some gestures is quite popular with a substantial amount of research already published on this topic (Cheng & Kuniyoshi, 2000; Kuzuoka, Oyama, Yamazaki, Suzuki, & Mitsuishi, 2000). Similarly, in this chapter the focus is to use a robot to portray conversational gestures, specifically if these gestures are to be performed by a tele-operator in a group tabletop interaction task.

One argument that has been put forward regarding the use of a robot as part of a tele-presence setting, where the robot is used to perform actions on behalf of a remote participant, is that the robot needs to provide humanlike gestures in order for the experience to be rewarding (Kose-Bagci, Ferrari, Dautenhahn, Syrdal, & Nehaniv, 2009). However, as proven before the human-likeness of these gestures need not be exactly like the ones being performed from a performer for them to be useful (Riek, et al., 2010). While, in the previous chapters, it was shown that humanoid robots can be used effectively to transfer pointing gestures from a remote location to the immediate vicinity of the robot, a greater set of gestures need to be evaluated for testing conversational gestures on a robot. Depending on the level of involvement, the topic and the capabilities and limitations of the people taking part in a conversation the type of gestures can be quite varied (Cohen, 1977; Efron, 1941; McNeill, Hand and mind: What gestures reveal about thought, 1992). As such there has been a number of attempts to classify the types of gestures people use in conversations (Efron, 1941; Ekman & Friesen, 1981; Freedman & Hoffman, 1967). According to Ekman and Friesen (1981), these gestures that people use during conversation can be of two types, illustrators, which are used as part of conversation, and emblems, which have a definite meaning and do not need to be supported by the conversation. In the context of having a robot as an embodiment for a remote participant, taking part in a group activity, a number of actions can be portrayed. While in the previous chapters, the actions portrayed were limited to just pointing gestures, in this chapter the gestures that are used are quite diverse but similarly to the previous chapters, it is important to evaluate the comprehension of the gestures portrayed. Since the number of gestures that a person can use during a conversation form quite a large gesture set, a sample gesture set, representing some conversational gestures, is going to be evaluated.

The selection of a suitable set of gestures, that can be used to evaluate the gesturing capabilities of the humanoid robot, is an important element of this gesture. Since gestures can vary quite a bit in meaning from one culture to another, it was important to find a set of gestures that would be able to be understood easily, and hence their comprehension assessed (Riek, et al., 2010). Taking inspiration from the work of Riek et al., a set of predefined arm gestures was selected (Riek, et al., 2010). The set of gestures used comprised of three gestures used by Riek at al. (beckon, give, shake hands), supplemented with gestures specific to a task based tabletop tele-presence interaction (point, turn). The gesture selected are all emblems, as defined by Ekman and Friesen (1981), although in this case they are being used without the accompanying audio. The gestures selected are described in Table 6.1-I.

Table 6.1-I: The Five Interactional Gestures

Gesture Name	Description
Beckon	The right arm is raised, palm upward, towards its face, makes two strokes, then the arm is retracted down.
Give	The closed right fist arm is raised to the centre of the torso, then the hand is extended, then the arm is retracted down.
Shake Hands	The right arm is raised, palm facing to the left up, and then fully forward, then the arm is retracted down.
Point	The right arm is raised, made into an index pointing, moved forward, then the arm is retracted down.
Turn	The right arm is raised, made into an open three finger grip, rotated round the lower arm, then the arm is retracted down.

It should be noted that the gestures listed in Table 6.1-I are limited to only one-handed gestures. Only one-handed gestures were used as two-handed gestures would require that there is high synchronicity between the two arms for joint gesture portrayals, which cannot be guaranteed on the current embodiment and also will allow for this type of gestures to be potentially implemented on a larger number of other platforms. These one-handed gestures can be decomposed in a number of parts or phases, depending on the model used these phases can be either three or five. The original model put forward by Kendon contained only three phases (Kendon, 1980), while the McNeill model possesses five (McNeill, 1992). Both models contain these sequential phases: preparation, stroke and retraction, with the McNeill model including holds phases in between the phases ending up with preparation, pre-stroke hold, stroke, post-stroke hold and retraction. In both models prior to performing the stroke, which is the most meaningful part of the gesture, the hand has to be put into position during the preparation phase and moved out of the position during the retraction phase. In McNeill's model, the hold phases allow the other phases to be clearly separated from the stroke phase and, by varying their length, they allow speech and gestures to be more easily synchronised.

It is to be noted that a pointing gesture is also included here, but it differs significantly from the pointing gestures from Chapter 4. In Chapter 4, the pointing gestures produced were generated autonomous by a specific scheme and was used for both portraying the gesture and also to designate specific objects. In this chapter, the pointing gesture is only used to portray the act of pointing and is produced from captured user motions rather than autonomously.

6.1.4 <u>Hypotheses</u>

As mentioned previously, there a number of different classification systems for gestures (Ekman & Friesen, 1981; Freedman & Hoffman, 1967), the classification systems however focussed on the purpose of the gesture. Thus while the same gesture can belong to different classes in different classification systems the purpose of them remain almost the same. A gesture does not only have purpose but also a descriptive form, this form of the gestures, used in this chapter, is presented in Table 6.1-I. One other element that gestures can have is what is referred by McNeill (1992) as the phases of the gesture. These phases are distinct elements of motion and can vary from two to three depending on the type of the motion. These distinct phases are important both in term of the appearance of the motion and also in terms of synchrony (Cassell, 2000). The approach that was selected for the portrayal of the gestures is expected to preserve part of the motion and maintain the position of the phases with respect to time. Since the main elements of the gesture are going to be preserved the following sub-hypothesis was formulated:

H 6.1 Scaled down gestures are as easily identifiable as full scale gestures from a humanoid robot avatar.

6.2 Method

Since the purpose of the experiment is to find out whether the control scheme allows for gestures to successfully modified, as a first step to the actual experiment it was important to find out whether the unmodified selected gestures, described in Table 6.1-I, were successfully identified. Therefore, a preliminary study was carried out using the humanoid robot, to evaluate how successful the implemented gestures were at being identified. Following from this preliminary study, the main study was carried out to evaluate the success rate of a scheme that is used to modify the gestures.

6.2.1 Participants

A total of 3 participants took part in the preliminary study (1 females, 2 males); aged 20 – 25 (M = 23, SD= 1.3). However, for the main study, a total of 12 participants took part in the experiment (3 females, 9 males); aged 22 – 32 (M = 28.1, SD = 3.93). It is to be noted, that 3 other participants took part in a preliminary study, which used unmodified gestures, were consequently excluded from the main study to prevent their prior knowledge to affect their judgment. A larger pool of participants was used with respect to the preliminary study as a higher variance in responses was expected when compared to the preliminary study.

6.2.2 Material

In order to test the proposed strategy, regarding human conversational gestures reproduced through robots, a suitable system was required. In this first subsection a description of the hardware configuration of the system used in this part of the investigation is provided. Figure 6.2-A provides a top-view of the implemented system. It is comprised of two different subsystems, the capture subsystem and the reproduction subsystem, denoted by the different boxes in Figure 6.2-A. Also, the remote user in the top left corner of Figure 6.2-A, had his actions pre-recorded and is not interacting live with the system.

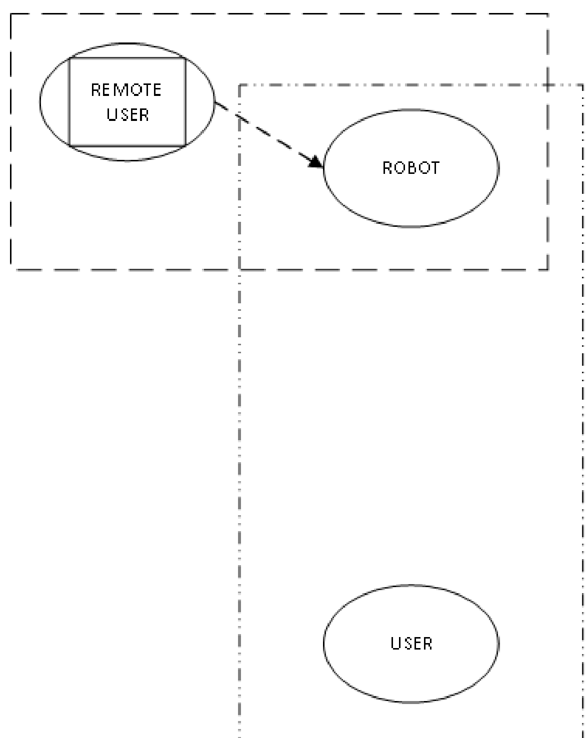

Figure 6.2-A: System Setup

To implement the system, a number of hardware components were required, apart from the robot, the humanoid RoboThespian (See section 3.2.1.1), that is used to portray the gestures of the remote user, a few other hardware components, making up the capture module, used to capture the performed gestures, were also required. In Figure 6.2-B, the capture module is portrayed with respect to its main components and with its link to the output components. The capture module is made up of two sub-modules, the body capture module and the hand capture module.

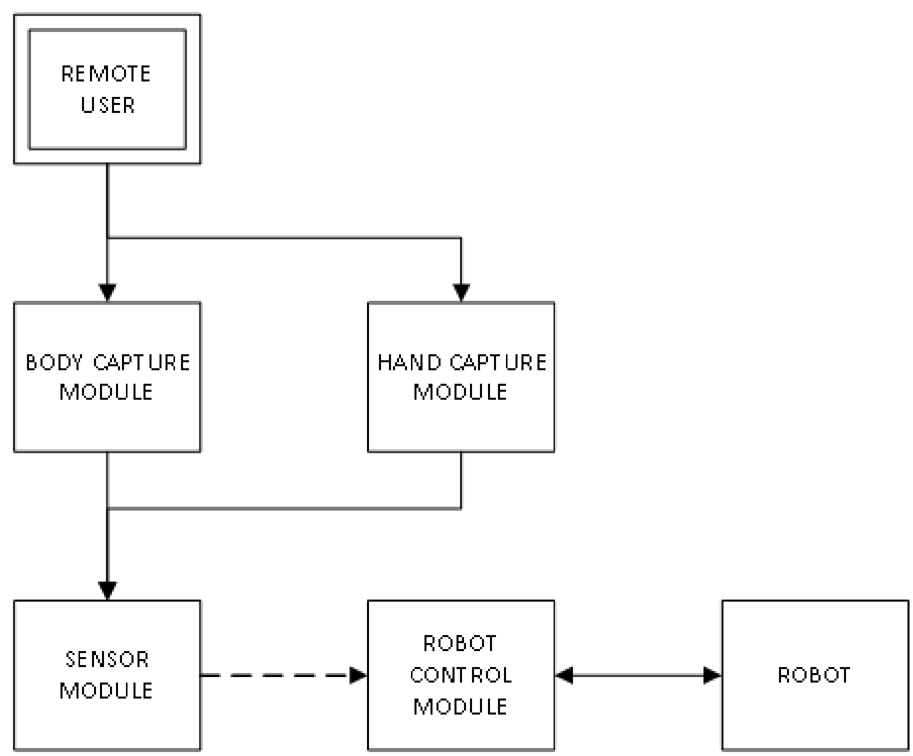

Figure 6.2-B: Capture and Output Components

For this system, the body capture module was implemented using the Kinect while the hand capture module was implemented using both the 5DT Data Glove and the Polhemus Patriot sensor; these additional hardware components are described below.

6.2.2.1 Kinect

6.2.2.1.1 Description

Figure 6.2-C: Microsoft Kinect

The capture component that was selected for the body pose capture system was the Kinect[17] (See Figure 6.2-C). The Kinect is a product originally developed as a component for the Xbox 360 console also developed by Microsoft. The Kinect has a number of subcomponents, an Infrared array projector, an RGB (Red, Green, and Blue) camera, an Infrared camera, a motor that allows the Kinect to be tilted and an array of microphones that can be used to capture audio. The Infrared projector and the Infrared camera operate together in order to create depth images, similar to a normal RGB image except that each pixel (individual part of the image) represents the distance from the camera rather than the colour component value. Both cameras can operate at a maximum of 30 frames (images) per second with a resolution of 640 by 480 pixels. The cameras also have a field of view of 60^{o} this is of particular importance with respect to the depth camera as while it has a range of 4 metres, it is only able to effectively capture depth data to a distance of 3.5 metres.

---

[17] Kinect® is a Product of Microsoft®. ***http://www.xbox.com/en-GB/kinect***

Moreover, the depth camera image describes an arc of view due to the limitation in range of the depth sensor (See Figure 6.2-D).

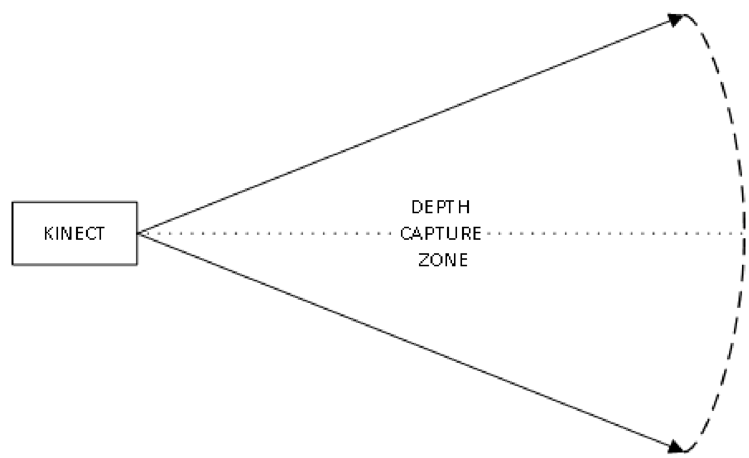

Figure 6.2-D: Kinect Depth Range

6.2.2.1.2 Kinect SDK

One important element of the Kinect is its SDK (software development kit) which contains, amongst other things the device drivers for the Kinect to be identified on devices running the Microsoft Windows operating system. The device drivers also allow the Kinect device to be used in more advanced ways. One of these ways is that the body position of an individual can be captured through the detection of certain key features of the skeleton.

It is to be noted that the SDK itself does not provide this information directly, but software applications built that incorporates elements of the SDK are able to access the information through streams of data. These applications are able to access not only the position of key elements of the body of a person but also the depth of most points in the image, allowing for the person that is being sensed to be able to interact with virtual objects.

6.2.2.1.3 Capabilities

Using the Kinect along with the SDK, two important data streams can be obtained: the skeleton data and the depth data. Using the Kinect and the SDK within a suitable environment, an environment without too much background Infrared sources or highly reflective surfaces, the Kinect can provide information on the physical setup of the environment as well as identify possible humanoid figures. The key advantage of the Kinect, in this task, is that it can provide relatively accurate skeletal and environmental setups without too much computational overhead. Also since it captures the body directly, it is easy to setup not requiring the participant to be wearing anything specific or calibrated for that matter. That is, allowing the capture of the unhindered motion from the participant.

6.2.2.1.4 Limitations

The Kinect suffers from a serious drawback where image resolution is concerned. This lack in resolution leads to two main problems when dealing with situations where the Kinect is used to capture pose data from a remote participant. The first of these problems is with respect to fine or subtle movement, while the second is the absence of certain key features for proper gesture capture.

The Kinect device is very useful to acquire the pose of a body or large movement but the restricted resolution mentioned above leads to fine movements being completely ignored by the sensor. This is an issue since fine movement can change the context of some gestures entirely, especially when it comes to hand gestures. Since the remote input capture device is meant to capture and reproduce the actions of a remote participant the Kinect alone is ill-equipped for that purpose.

While the skeleton model provided by the Kinect provides information about the main parts of the body, it however lacks information with respect to the hands. As mentioned previously the use of hands plays an important part in gesturing and whole categories of gestures are impossible to reproduce without information about the hands. The best that the Kinect can offer in terms of hand information is an approximation of where the centre of the palms may be. This type of information is not very useful and therefore an alternative, to using the hand information from the Kinect, was devised.

6.2.2.1.5 Conclusion

The data streams obtained from the Kinect allows the action and gestures from a person to be captured. The data is also obtained directly from the subject without him/her needing to wear anything that may interfere with the actual gesture that he/she is attempting to perform. This is in sharp contrast to other gesture capture system such as the Vicon capture system which requires the subject whose movement is being captured to wear specific tags or markers on his/her body for the movements to be captured. Wearing these markers has the possibility of constraining and affecting the behaviour of the individual and thus limiting the range of motion. Concerning the accuracy of the system, it is sufficient for the task when it is being supplemented with a fine-gesture capture system. Having access to the depth data also allows more information to be extracted from the environment, this will be further elaborated in Chapter 7.

6.2.2.2 5DT Data Glove

Figure 6.2-E: 5DT Data Glove 14 Ultra

Another input capture component that was used was a 5DT Data Glove 14 Ultra®[18] (See Figure 6.2-E). While the previous input capture component was more oriented towards capturing the overall body position of the remote person, a device was also required to capture finer input from the hands since a lot of human gestures require the use of hands and fingers. The data glove device in this section is similar to a normal Lycra fingerless glove. It however contains sensors that can detect deformation in the material enabling the hand gestures to be captured.

---

[18] 5DT Data Glove 14 Ultra ® is a Product of 5DT (Fifth Dimension Technologies Inc. ***http://www.5dt.com/products/pdataglove14.html***

6.2.2.3 Polhemus Patriot Sensor

The data glove described in the previous section was able to capture the hand gestures which included the state of each of the fingers, being either open or closed, and the separation between the fingers. However, using only the data obtained from the glove would not be sufficient to accurately portray the gestures being performed by the remote user as radial rotation and wrist flexion is not available. It is in order to capture that information that the orientation sensor is required.

Figure 6.2-F: Patriot Sensor

The orientation sensor that was selected was the Polhemus Patriot®[19] (See Figure 6.2-F). This sensor is actually a 3D positioning sensor, providing position (X, Y, Z) and orientation (Yaw, Pitch, Roll) values to the millimetre and degree respectively. It is comprised of three components: the emitter, two interchangeable receivers and the interface device. It can be used to obtain both the position and the orientation of the receiver with respect to the emitter and in the coordinate frame that the emitter has been previously set. This sensor uses magnetic fields to calculate the position

[19] Polhemus Patriot is a Product of Polhemus. ***http://polhemus.com/motion-tracking/all-trackers/patriot/***

information of the receiver with respect to the emitter and it can then transform that data into real world coordinates. As mentioned above, this sensor provides two types of data, position data and orientation data, however only the orientation data is used by the system described here. This approach was selected to be in line with the rest of the system which focusses on orientation and relative positions to other points rather than absolute positions. This approach provides the advantage of being scale independent, i.e., that the captured data can be easily ported to a 2 foot or a 6 foot humanoid robot.

6.2.2.4 Gesture Reproduction

The scheme used for gesture capture and reproduction in both the preliminary study (See section 6.2.3.1) and the main study (See section 6.2.3), is described here; though in the main study the gestures are also modified. This scheme is made up of four distinct phases: Capture, Process, Transform, and Execute. While the other three phases remained the same in both studies, the Transform phase changed quite considerably between the two studies, more on the modification performed in the Transform phase is elaborated in section 6.2.2.5.

In the first phase of the scheme, the Capture phase, the sensor modules together with its sub-modules and the related hardware are used to capture data from the body of the performer. Using the Kinect device and its related module, positions of the main parts of the main joints of the arm (shoulder, elbow, and wrist) relative to the position of the Kinect, are obtained. For data of the configuration of body parts under the wrists, hand orientation and configuration of the fingers, the Patriot sensor and data glove are used. This data is captured at 30 frames per second, the maximum capture rate of the Kinect device.

In the second phase of the scheme, the Process phase, the raw data from the different devices are processed into useful information for controlling the robot. As part of this phase, the data from the different devices are integrated into full arm frames for the robot. Using the data from the Kinect module, data about the position of the different joints of the arms with respect to the position of the Kinect, unit vectors of the different of the orientation of the different parts of the arm, upper and lower arms, are produced. The unit vectors for the upper and lower arms are obtained by subtracting the positions of the elbow and the wrist from that of the shoulder and the elbow respectively and the vectors obtained are then normalised. Using unit vectors instead of actual positions for the parts of the arm allows for the gesture data to be transferred to embodiments that do not match exactly the dimensions of the source of the gesture. This allows for auto-scaling of the gesture with respect to the size of the embodiment, which result in a similar visual appearance of the gesture irrespective of the scale of the embodiment.

While the unit vectors for the direction of orientation is adequate for the main parts of the arm, they are not enough for the hand orientation, since information on the forearm rotation is lost. A different strategy is therefore used for the hand orientation; it involves using information from the orientation sensor to produce two unit vectors. The first of these vectors is the relative position of the first knuckle of the middle finger from the wrist and the second is that of the first knuckle of the fore finger from that of the little finger. Correct positioning of the orientation sensor is required in order to obtain the necessary information and so it is fixed on the back of the hand at the base of the first knuckle of the middle finger. These vectors are perpendicular to one another and help obtain the full orientation of the hand. The process of obtaining the new vectors requires the base vectors to be multiplied by the transformation matrix obtained from the orientation data as per Equation 6.2-i.

Equation 6.2-i: Vector Transformations

$$\begin{bmatrix} x_b \\ y_b \\ z_b \end{bmatrix} \quad \begin{bmatrix} \cos\varphi\cos\psi + \sin\varphi\sin\theta\sin\psi & \cos\theta\sin\psi & \cos\varphi\sin\theta\sin\psi - \sin\varphi\cos\psi \\ \sin\varphi\sin\theta\cos\psi - \cos\varphi\sin\psi & \cos\theta\cos\psi & \cos\varphi\sin\theta\cos\psi + \sin\varphi\sin\psi \\ \sin\varphi\cos\theta & -\sin\theta & \cos\varphi\cos\theta \end{bmatrix} = \begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix}$$

In Equation 6.2-i, x_b, y_b and z_b represent the coordinates of the base vector while x_n, y_n and z_n represent the coordinates for the transformed vector and while φ, θ and ψ represent the values of yaw, pitch and roll respectively.

Also, the configuration of the finger states are obtained directly from the gesture data stream of the data glove. Therefore, each of the pose frames obtained is made up of a series of unit vectors representing the orientation of each of the arm links (upper arm and forearm), with two more for the orientation of the hand, and data about the bend state of each finger.

In the third phase of the process, the data is transformed into the robot's coordinate frame, and used to generate a set of joint angles that can best approximate the link orientations. Using the demand orientation vectors and the current orientation vectors of the corresponding arm link together with Equation 6.2-ii and Equation 6.2-iii, the required change in angle can be obtained.

Equation 6.2-ii: Component Angles

$$\propto_h = \tan^{-1}\frac{x}{z}, \propto_v = \sin^{-1} y$$

In Equation 6.2-ii, $\propto_h$ and $\propto_v$ are the horizontal and vertical components of the angle and x, y and z are the coordinates of the vector.

Equation 6.2-iii: Angle between vectors

$$\cos \propto = \frac{a.b}{|a|\,|b|}$$

This process is straightforward when both the demand and current vectors lie on the plane, defined by the point of rotation and the axis of rotation, however if this is not the case, projection of the vectors onto the plane is required, as discussed in section 4.2.2.5. Each of the demand angles is tested for values exceeding the range of the joint and if they do then the value of the closest boundary is selected, else the actual demand value is selected. Considering that the range of the joints are restricted, an updated transformation matrix is also used in the processing part to keep track of the current transformations in the coordinate frame and allow for the lower joints to more closely match the original gesture. It is during this phase that the modification of the gestures is carried out in the main study.

Finally in the execute phase, the computed commands are sent to the robot for it to produce the captured movement. In the case where the robot receives a command for a joint that is currently moving, the value from the command is used to overwrite the demand value for that joint and a modified trajectory for that joint is produced without requiring the joint to be halted.

6.2.2.5 Gesture Modification

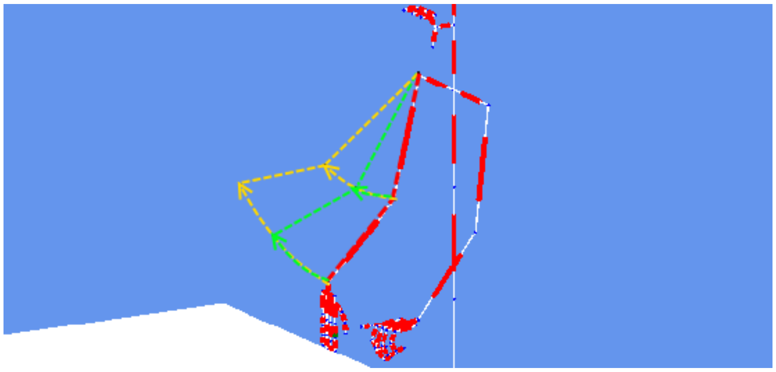

Figure 6.2-G: Modified Gesture

The approach, that is used to modify the gestures in order to enable speech alignment, is for the gestures to be scaled down during each of the phases of motion. This approach decreases the range of the motions across each phase of the gesture while still preserving the presence of each phase, i.e., a uniform scaling down of the gesture across each phase is carried out instead of just focussing on specific phases. Using this approach all the phases of the gestures, as described by Kendon, are still present, but with the overall motion reduced. With this approach all of the required elements for the gesture to be identified are still present in the modified gesture though the motion is more subtle (Kendon, 2004). Moreover, this approach is computationally cheap and therefore creates no significant additional delay in the gesture reproduction process, which could be incurred if phase analysis and other scaling methods were used instead. It is to be noted that certain elements of the gestures were purposefully not scaled down, mainly the finger configuration since the fingers operate in a binary state (open or closed) and scaling them down will cause the fingers to be unable to portray any gesture. Another approach to this one would involve the use of predictive models, but since a computationally cheap approach is needed and that accuracy of the prediction cannot be confirmed, which could lead to the wrong gesture being predicted, this one was favoured and predictive models are beyond the scope of the thesis.

This approach is selected in order to allow a robot, which moves at a slower speed than a human, to be able to reproduce gestures that are as close as possible to being synchronized with the speech from the tele-operator. It does not implement any gesture recognition, with the gesture scaling being triggered during speech so that all movement that occurs while the tele-operator is speaking are scaled. As mentioned before, for each command frame of joint angles that is generated, the angles of the arm joints are reduced by a fixed scaling factor.

By comparing the movement times for different gestures being performed by a human, with the equivalent movements performed by the RoboThespian, it was observed that the robot is able to move at about half (average factor of 0.57) the speed of a human for typical gestures, limiting the speed of the hand to a maximum of 0.5 m/s when the movement is being performed from the elbow joint. Hence, a joint scaling factor of 0.5 is used to investigate the validity of producing recognisable gestures at a smaller scale. In Figure 6.2-G the range of an example gesture, handshake (described in Table 6.1-I.) is shown with respect to both the original range (yellow) and the modified range (green).

6.2.3 Procedure

From the conclusion drawn from the preliminary study, the tele-operation control method has been shown to be able to correctly reproduce the gestures from Table 6.1-I. Since the non-scaled gestures have been successfully reproduced, the main study was conducted to find out if the scaled gestures were as comprehensible as the non-scaled ones.

Using a similar setup as with the preliminary study, i.e., the same humanoid robot and the same set of gestures, listed in Table 6.1-I, the main study was conducted with the participants having to identify the scaled gesture reproductions, instead of the non-scaled gesture ones. The scaled gesture reproductions were produced, using the gesture scaling method described in section 6.2.2.5, from the same pre-recorded set of gestures that have been correctly identified in the preliminary study.

The order of the presentation of the scaled gesture reproductions was varied and a table of all possible combinations for the order of the gestures was produced, from which the order of the gestures for the presentation for each participant was selected. Once the presentation of the gestures to the participants was over the particular sequence was removed from the table. This was carried out in order to allow the participants to each see a unique sequence and so eliminate any ordering effect.

6.2.3.1 Preliminary Study

For the preliminary study a capture system was used to produce recordings of each of the gestures. Although the capture system was primarily designed to operate in real-time, the recordings were created in order to allow each participant who took part in this study to experience the same set of gestures. Each participant was tasked during this study to identify each of the gesture from Table 6.1-I and was therefore given a questionnaire (See Appendix A.1) to fill in with what action was being portrayed during each gesture. The questionnaire used open-ended questions since the gestures being performed by the robot were the non-scaled gestures. The presentation order of the gestures was varied between participants. The participants were standing approximately 1.5 metres away and directly in front of the robot during this study.

All participants successfully identified each of the gestures being performed by the robot. A Krippendorff's α value of 1 reflected this perfect agreement. Hence, it can be concluded that the tele-operation control method was able to correctly reproduce the selected gestures.

6.2.3.2 Task

Prior to running the experiment, the participants were instructed in the task that they were to perform, that is to identify the gesture being performed from the humanoid robot. They were also provided with a questionnaire (See Appendix A.2) that served a similar purpose as in the preliminary study; however, the questionnaire in this study was made up of multiple choice questions. In order to prevent any process of elimination effect that can arise from the use of multiple choice questions, the options for the identifiable gestures was increased to include 5 more gestures than the 5 targets. These additional options were from the same class of one handed communicative gestures. The Table 6.2-I, contains a brief description of each of the gesture options in the questionnaire, note that these descriptions were not given to the participants.

Table 6.2-I: The Set of Possible Gestures

Gesture Name	Description
Beckon	The right arm is raised, palm upward, towards the face, makes two strokes, then the arm is retracted down.
Good	The right arm is raised to the centre of the torso, hand in a fist with the thumb sticking out upwards, then the arm is retracted down.
Give	The right arm is raised to the centre of the torso, hand in a fist, then the hand is extended, then the arm is retracted down.
Okay	The right hand is raised up to the upper torso with the thumb and fore finger in a pinch like fashion with the other three fingers extended, then the arm is retracted down.
Phone	The right hand is raised to the head with the thumb and little finger extended while the other are closed and then the arm is lowered.
Point	The right arm is raised, made into an index pointing, moved forward, then the arm is retracted down.
Shake Hands	The right arm is raised, palm facing to the left up, and then fully forward, then the arm is retracted down.
Success	The right arm is raised to the upper torso, palm facing forward with the fore and middle finger extended and apart with the other fingers closed, then the arm is retracted down.
Turn	The right arm is raised, made into an open three finger grip, rotated round the lower arm, then the arm is retracted down.
Wait	The right arm is raised, palm facing forward, and moved forward, then the arm is retracted down.

In order to decrease the cognitive load of the participants, they were allowed to fill in the questionnaire after each modified gesture was performed. Finally, since the participants in the experiment came from different backgrounds, both in terms of countries and culture including non-native English speakers, they were asked prior to the conducting of the experiment if they understood what each of the gestures from the list meant. If they did not understand a specific gesture a situation where the gesture would be used was described without actually portraying the gesture. This was only contingent on the participant admitting their confusion to the meaning of the gesture and only occurred twice, specifically with the ‘Okay’ gesture.

6.3 Results

The data obtained from the modified gesture recognition user study was analysed and Figure 6.3-A was produced. In Figure 6.3-A, the portrayed gestures are set against the user input with the correct performance and the chance outcomes being represented by the two horizontal lines with, the top and bottom one respectively. It can be observed that the modified gestures are easily identifiable, with a majority of participants correctly identifying all gestures. It is to be noted that this was obtained using participants that had not previously seen the gestures from the robot, or from the performer. This would suggest that the gestures that were modified in order to be in line with their time slot were still recognizable by the participants.

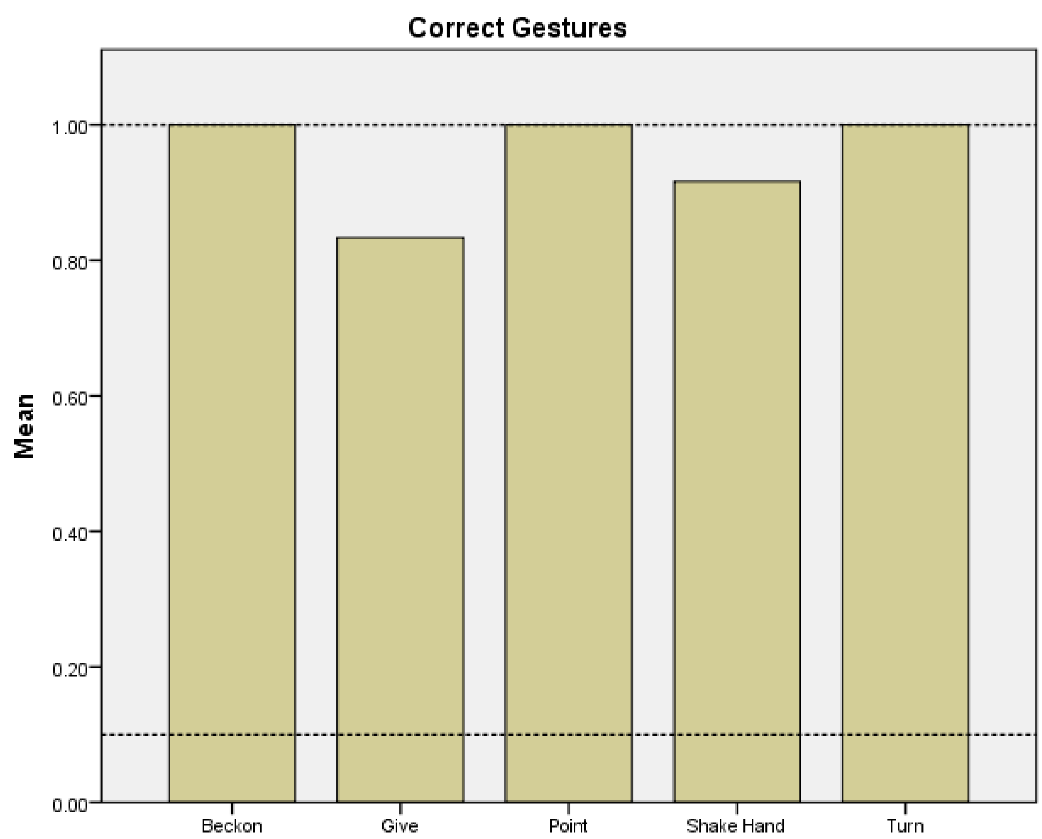

Figure 6.3-A: Correct Gestures

A statistical analysis was performed on the data to see if there was a relationship between the portrayed gestures and the selected gestures. Since the data being analysed is of nominal type and possesses more than one category a Pearson's chi-square test was used to test this. The results of the test showed a strong statistical significant relation, χ^2 (24, N = 60) = 240, p < .001, between the portrayed gestures and the selected gestures. Moreover the result of a Cramer's V test, selected due to more than two categories in the data sets, showed a very strong relationship between the two variables, $\phi = 1$, p < .001. Finally the result of agreement test using Cohen's Kappa, showed strong agreement between the portrayed gestures and selected gestures, K = .938, p < .001. From these results it can be seen that the selected gestures from the participants are significantly similar to the portrayed gestures by the robot using the gestures modification scheme.

In order to determine the success of the movement scaling in aligning the gestures with speech, the differences in performance times for the human and robot performed gestures were analysed. According to Kendon (2004), a gesture is made up of 3 distinct phases: preparation, stroke and retraction, with the stroke phase being the most effortful part of the gesture that accompanies the element(s) of the speech the gesture is associated with (Kendon, 2004). The preparation phase is moving the hands into position to start the stroke phase. Both these phases are vital to the gesture being synchronised with speech. Hence, for each gesture the preparation and stroke phases were identified and the durations were recorded, with respect to both the human and the reproduction on the robot. These findings are shown in Table 6.3-I.

*Table 6.3-I**:** The Time Values (in ms) for the Five Interactional Gestures*

Gesture	Human Stroke Duration	Robot Stroke Duration	Human Preparation Duration	Robot Preparation Duration
Beckon	3620	3761	582	561
Give	2982	3153	642	650
Shake Hands	1571	1476	555	453
Point	2039	2208	552	440
Turn	2142	2074	610	550

An analysis of the time taken for the motion of the robot and the human for the same gestures with respect to the constituent phases of the gestures was carried out using a dependent T-Test. The normality of the distribution of the different samples was tested together with the presence of any outliers. Since there was no outliers and the samples were approximately normally distributed the dependent T-Test was carried out first for the preparation phase of the gestures on both the human (M = 2471, SD = 819) and robot (M = 2534, SD = 408) execution times. The results of the test showed no significant difference between the two times, t (4) = 1.067, p = .346.

When the dependent T-Test was carried out on the stoke phase of the gestures on both the human (M = 588, SD = 17.1) and robot (M = 531, SD = 38.6) execution times, a similar result was obtained whereby there was no significant difference between the two execution times, t (4) = 2.496, p = .067.

Since the duration of the stroke phase was found to be not significantly different between the two embodiments, it can be concluded that the stroke phase, the most meaningful part of the modified gestures from the robot will last approximately as long as the stroke of the source gesture. Another important element of the gesture is the preparation phase of the gesture, for this one also it was shown that there was no significant difference between the duration from the different embodiments. Therefore, it can be concluded that the modified gestures will be in approximate synchronisation with the original gestures, since no significant differences was found between the duration of the corresponding phases between the original and modified gestures.

6.4 Discussion

In the first part of the experiment the participants were tasked with identifying the different gestures being portrayed by the humanoid robot. Each participant was shown five different gestures and was asked to identify each one from a series of ten possible options. From the results obtained it was shown that out of the five gestures, the participants were able to correctly identify four of the five gestures with a majority able to identify all five gestures correctly. After a careful analysis of the results with a focus on the errors in the results, it was found that not all three errors were on the same misidentified gesture. Of the three errors, one was misidentified as a Success gesture while being in fact a Turn gesture, while the other two errors were Give gestures being misidentified as Good gestures.

For the first error, where the Turn gesture was misidentified as a Success gesture, it may be due to the similarity between the configurations of the hand of the robot for each gesture. Taking only the state of the fingers into the consideration, both share a similar state with the fore and middle fingers of the hand open while the other fingers are closed. The main difference between the two is the rotational motion of the wrist in the Turn gesture. It is therefore possible that the participant may have focussed solely on the fingers configuration and overlooked the rotation motion and therefore confused the two gestures. The second error in the results occurred twice since both the other errors were on the same misidentification, that is, Give gestures misidentified as Good gestures. The reason for this error may lie in the manner in which these actions are performed by the robot. Good gestures, when performed by a human, involve the thumb sticking out and pointing upwards while the other fingers are closed into a fist raised to the torso level. The Give gestures involve the use of a closed fist raised to mid torso and then push forward and with the fingers opening as in a drop motion. On the robot however, the thumb is not actuated leaving the thumb perpetually in a half open position. The state of the thumb could account for the misidentification of the gesture.

From the results of the first part of the experiment, it can be observed that the gesture modification scheme was successful in conveying the gestures that were captured from the performer. While there were some instances of gestures from the robot where the wrong gesture was identified, they occurred in a small percentage of cases. After quantitatively analysing the data it was found that the errors did not provide enough divergence from a perfect score to be significant, thus supporting the hypothesis H 6.1. Furthermore, these errors happened when the gestures were being identified on their own, that is, only the motion was performed with no accompanying speech. With the high success rate and accompanying speech as it is intended to be used as part of the interaction process, it is likely to be even more successful and for the error rate to be decreased.

For the second part of the experiment the data is summarised in Table 6.3-I. It can be observed that the values of human and robot gesture phases across each gesture are fairly close to one another. Comparative analysis found that the maximum difference between the human and the robot for the stroke phase was 171 ms, and the maximum difference between preparation phases was 112 ms. Since the stroke phase follows the preparation phase, it can be calculated that if the gesture on both embodiments are started at the same time the stroke phase lag between the two embodiments will be a maximum of 112 ms (for the examined gestures). Taking into account the values obtained for the duration of the stroke phase (averaging 2534 ms see Table 6.3-I) and the size of the movement involved a lag of 112 ms is unlikely to be perceptible.

However, the robot is not able to move immediately after receiving the command, there is an execution delay on the robot upon receiving the command. This execution delay has been estimated to be variable with a maximum value of 250 ms from the sending of the command. This execution delay is due to the architecture of the control system of the robot which operates using a time slot system, the commands received within a timeslot are scheduled for execution in the following time slot. Each timeslot is approximately 250ms. Therefore depending, at what point in the timeslot the command is received, the execution delay can be up to 250ms. When the execution delay is combined with the preparation lag, the stroke phase on the robot is delayed by about 375 ms from when a human would usually start the stroke part of the gesture. This value is not ideal for use in conjunction with real-time audio communication, where a lag is expected to be less than 300ms (Bohren, et al., 2013).

One possible option that can be implemented when these gestures are to be used as part of a tele-presence experience is to add a delay to the audio stream. This delay can be set to either the full 375 ms, mentioned earlier, or just by 250 ms to allow both the gesture and the speech to be in synchrony, this is not problematic, since even with a 375 ms delay, the introduced audio delay would still be within the 400 ms limit for VoIP set by the International Telecommunication Union (2003). Hence, for the range of gestures tested here, comprehensible gestures synchronised with speech could be produced using the described system, with acceptable lag in communication.

6.4.1 Summary

In this part of the investigation, it was shown that the use of the implemented gesture modification scheme was successful at producing recognisable gestures. The participants that took part in the experiment were able to identify the gestures with a high degree of success. This high success rate was achieved despite the fact that there was no accompanying speech to provide context to the gesture motion. Furthermore, the use of the gestures as part of a tele-presence experience where the gestures are supported with corroborating words, is likely to yield even better results. This will be further explored in the following chapter.

The implemented gesture transformation scheme is critical for a robot avatar both as a method to allow implicitly captured gestures from a performer to be executed on a humanoid robot; and also to allow these gestures to be performed in a time sensitive context. While the robot was able to execute the unmodified gesture, it was not able to do so in a timely manner since it was only able to move its limbs at about half the speed used by the performer. Without such a scheme, the delayed gesture being performed by the robot could lead to misunderstanding or confusion from an observer since the gestures may occur out of context especially if being used as part of a tele-presence experience. The proposed scheme involved the uniform scaling of the movements across every phase of the gesture and for every joint involved in the gesture. The success of this scheme while limited to one arm gestures, has some useful implications for the design of robots in general. In order to move the mass of the arm at high speed, powerful motors are usually required. The results presented here suggest that this is not essential for robots designed to produce multi-modal communication as a smaller range of motion and lower speed are adequate. On a similar note, the use of this scheme can also be extended to non-humanoid robots to portray similar gestures at least as far as end effector trajectory, joint coordination and link orientations is concerned. Finally, it is to be noted that one crucial element for the use of this scheme on non-humanoid robots is the design of the end effector of the robot.

The selected gestures for the experiments, limited to one arm communicative gestures, showed that the scheme was quite successful at implementing these gestures within the time constraints of the context, that is, in synchrony with the accompanying speech. It should be noted that while the selected gestures were successfully recognised, some gestures require more than identification to be useful. For example pointing gestures, apart from being correctly identified, also have to be pointing to the correct objects or areas for them to be useful. While the pointing gesture obtained from the proposed scheme is recognisable, it does not however transfer the additional information about which object is being targeted. So while the scheme produces a recognisable pointing gesture the limitation of range on the modified gesture is likely to prevent the targeted object to be correctly identified. This will be an issue that will be addressed in Chapter 7.

6.5 Chapter Conclusion

In this chapter, a gesture modification scheme that would allow a humanoid robot that moves at subpar human levels in terms of both range and speed was investigated. It was implemented through the RoboThespian robot and attempted to produce comprehensible gestures from gestures captured by motion capture system. Participants were able to correctly identify the set of gestures presented, even when they were scaled down; the purpose of which being the alignment of the scaled gesture with co-occurring speech. Hence, with the implemented gesture modification scheme, the robot can now be used to portray a more extended set of gesture. Although the context for the use of the robot is to use it in a remote tele-operation setting, a similar setup as described in this chapter is going to be used. This is going to be carried out in the following chapter where the previously mentioned limitation with respect to pointing gestures will be addressed to ensure that the targets of the deictic gestures are correctly identified. Finally, as alluded in this chapter, the scheme developed for the humanoid robot is going to be modified and implemented on a non-humanoid robot for the same purpose.

Chapter 7 Robot Presence

7.1 Chapter Introduction

In the previous chapters, the main focus has been on the gesture aspect of interaction which included methods by which robotic embodiments acting on behalf of a remotely located operator would be able to create or copy gestures from the operator. This was with respect to the context where a geographically distributed group of people would be able to collaborate in a way that would be similar to the group members being co-located. This focus was both on the development of methods to enable appropriate portrayal of gestures for the task and the impact of having a robotic embodiment acting on behalf of a remotely located person on the dynamics of the group. In this chapter, the focus is on the impact, of using robot avatars in such settings, on the level of presence that the avatars have on the group collaboration and dynamics. Presence, being an important element in this thesis, prior to investigating it, it was necessary to find appropriate methods by which the avatar enabled presence could be accurately exhibited. From previous works, it has been shown that gestures are a contributing element of presence (Adalgeirsson & Breazeal, 2010; Takayama, Groom, & Nass, 2009) and consequently implementing gestures became a requirement, prior to investigating presence. It is to be noted that while presence has not been investigated in previous chapters, some aspects of it may have contributed to the results obtained and conclusions drawn in the previous chapters. Since the implemented gestures use schemes developed from the previous chapters, in this chapter some issues of the schemes are also addressed. One such issue was identified in Chapter 6 with respect to pointing gestures; unlike the other gestures used in that chapter, pointing gestures have an additional requirement beyond identification, the target of the pointing. That is, the object being designated by the pointing gesture is as important as the proper hand and arm configuration for the gesture. This involves both the etic, the position of the hand and fingers, and the emic, the configuration of the hand, of the gesture as defined by Wilkins (2003). One focus of this chapter is how to successfully integrate the two previously developed schemes into one that would allow pointing gestures and other modified gestures to work in concert. Furthermore, since the main motivation behind this thesis is the enhancing of the tele-presence experience while using robotic embodiments as avatars, it is within the scope of robotic presence that the scheme is going to be evaluated. This will involve the use of the integrated scheme on different types of robot, a humanoid robot and a non-humanoid robot and evaluating its impact on the level of presence being exhibited by the local robotic embodiment of the remote operator.

7.1.1 Regarding Presence

As previously mentioned, the concept of presence is usually used with respect to entities that share the same physical environment and, according to Steuer, their perception of it (Steuer, 1992). While the use of this concept is often with respect to a physical tangible environment, it is not limited to such an environment as presence has been successfully established in non-physical environments (Brown & Bell, 2004). Considering that the term presence is usually defined with respect to the state of being physically in a location, the other types of non-physical presence are usually preceded by the context of this presence (Lombard & Jones, 2006). Therefore, in the case of the presence occurring in virtual environments, the term virtual presence is usually used to denote that type of presence and has been the subject of research into online social interaction (Brown & Bell, 2004). On the same note, presence can also be extended, through the use of technology, to other physical environments where the user is not physically present (Kidd & Breazeal, 2004). This remote presence or tele-presence can be achieved when a person, remotely located from the local environment, can either perceive the environment, provide a presence via an embodiment, and/or affect the environment (Kidd & Breazeal, 2004). While this can be achieved, to different degrees, using either static or dynamic robots (Adalgeirsson & Breazeal, 2010), other types of technology, one of which makes use of holograms, have also been proven to be successful (Blanche, et al., 2010). Finally, one of the most basic forms of tele-presence that is commonly available is the use of video conferencing applications that allow groups of people to collaborate over long distances (Augestad & Lindsetmo, 2009).

One term that is often associated with the use of interaction in non-co-located physical settings, is the term avatar. While the term avatar has been previously defined, as a representation of a being in an environment that the being is not physically present in, in the context of tele-presence it is more specifically a place-holder object or robot which may also act as an interface (Adalgeirsson & Breazeal, 2010; Tsui, Desai, Yanco, & Uhlik, 2011; Coradeschi, et al., 2011). Whereby, the avatar not only allows the actions of the remote person to become visible to the co-located collaborators , but may also, in certain situations, be the object to which queries are addressed to when addressing the remote person (Adalgeirsson & Breazeal, 2010).

Since the avatar is tasked with representing the actions of the remote person, on whose behalf it is acting, it is not uncommon for avatars to display attributes of the remote person, this can be done in a number of ways, two of which being direct image representation (Buxton, 1992) and through the use of representative gestures (Seo, Park, & Yang, 2003). In this thesis, the avatars that were used to embody remote participants have made use of robotic embodiments with added focus on the use of gestures from the robot avatars to express the actions of the remote participant in group interaction settings. It is to be noted that since the focus here is on how the use of gestures may enhance the experience, careful steps were taken to only represent the remote participant through these embodiments, acting on behalf of the remote person. Consequently, it is only through gestures and audio that the local participants can perceive the remote participants' states and not through any actual image representation of the remote participants.

Having limited the modes of interaction between the participants to gestures and audio, the impact of the use of robotic embodiments with implemented gestures from the remote participant on the exhibited presence can be assessed. While taking into account the results from the previous chapter, humanoids using the previously described gesture production schemes are expected to exhibit a higher level of presence than non-humanoid using their default gesture production schemes. However, since the concept of presence is multi-dimensional and there are several subtypes of presence that can be exhibited concurrently, an appropriate model for the presence to be assessed was required (Garrison, Anderson, & Archer, 2001). In the work of Garrison, Anderson, and Archer, which dealt with presence in a setting where computer based conferencing was used for distance learning purposes, the context for tele-presence established there shares similarities with the type of implementation of presence used in this thesis (Garrison, Anderson, & Archer, 2001). Furthermore, the model that was used subdivided the tele-presence experience into three dimensions of presence: physical, cognitive and social (Garrison, Anderson, & Archer, 2001).

The first of these, physical presence, is the one that most closely matches the traditional definition of presence, being located in a specific environment. However, exhibiting a physical presence in the context of tele-presence extends the definition of presence by adding aspects of interaction with real-world objects. This interaction can be either direct, which involves manipulation of the objects, or indirect, referring to but without actual manipulation (Kuzuoka, Oyama, Yamazaki, Suzuki, & Mitsuishi, 2000; Rosenberg, 1995). As shown earlier in this thesis, with respect to the level of physical presence, the use of robot embodiments provides significant gains in performance when compared to settings without any physical embodiments. It should be noted that a level of physical presence was also exhibited by the robots when they were not being tele-operated by the remote operator and so the level of the physical presence of the embodiment is a compounding of both the robot and that of the remote operator. It should also be noted that the use of robotic technology can also add a further level of physical presence as the remote person can also feel the object in the local environment using haptic feedback technology (Roke, Spiers, Pipe, & Melhuish, 2013).

Another type of presence according to the above model is the cognitive presence. Contrary to the previous type of presence, this type of presence can only be exhibited by an entity as it involves the communication between two or more entities in order to create understanding of the situation (Garrison, 2006). So, while robotic embodiments can be used to provide the vector through which this presence is exhibited, they do not exhibit it by themselves although, as shown earlier in this thesis, the use of a robot together with an appropriate algorithm can definitely have an impact on the level of presence provided. This was established earlier where the use of robotic embodiments performing meaningful gestures provided measureable gains in terms of group collaboration which led to a higher level of effectiveness in performing the assigned task.

However, the range of supported gestures in these earlier group experiments was rather limited especially with respect to the last type of presence in this model. This type of presence, social presence, can be defined as how an individual can perceive the other members of a group on an interpersonal level (Biocca, Harms, & Gregg, 2001) and as such has had a lot of interest with respect to online communities (Aragon, 2003; Stacey, 2002). One of its subtopics of interest is the role that the medium of communication plays in the overall level of social presence perceived (Garrison, Anderson, & Archer, 2001). According to the social presence theory (Short, Williams, & Christie, 1976) all the different methods of communication allow a certain level of social presence with the ones that allows for more interpersonal collaboration exhibiting a higher level of social presence. Therefore, a lack of representation of the actions of the remote operator can have a negative impact on the social presence of the remote operator as demonstrated by Adalgeirsson and Breazeal (2010). Conversely, it is expected that the use of robot embodiments with added gesture support, where the gestures are local representations of the gestures of the remote operator would allow for higher levels of social presence.

On the topic of social presence with respect to robot embodiments and gestures, two important works are Takayama et al. (2009) and Adalgeirsson and Breazeal (2010). Interestingly enough, contrary to the normal definition of social presence which test the social presence between individuals, in the work of Takayama et al., the social presence exhibited by a humanoid robot toy used a set of specific pre-programmed behaviours (Takayama, Groom, & Nass, 2009). While the robot did not have action autonomy, that is, it was tele-operated by operators, it was perceived as the sole initiator of the presence and the operators did not interact directly with the participants. The action of the remote operators, in that example, were limited to the selection of the behaviour to be exhibited by the robot. Finally, despite the fact that the source of the presence was a robot with a limited gesture set, there was a definite impact on the social presence.

On the same topic of robotic gestures and social presence, but with a different implementation is the work of Adalgeirsson and Breazeal, where a non-humanoid robot was used to test the overall presence in a group interaction activity (Adalgeirsson & Breazeal, 2010). The robot used in that study was a mobile platform with some human features in the form of two small arms to depict the gestures of a remote operator and a screen to allow face-to-face interaction. Contrary to the previous example, it was the remote operator that was the initiator of the interaction with the participants, not the robot embodiment. As such, during that study, the operator used the facial and arm gestures, which were captured using a rig, to establish its presence through the robot. From this study it was found that transferring gestures from an operator played an important role on the level of presence exhibited. Consequently, it is expected that the robot that behaves more similarly to a human being is going to be able to exhibit a higher level of presence and so the humanoid robot using the scheme is expected to exhibit a higher presence.

From the analysis of these work, some important elements were found with respect to the use of gestures from robotic avatars for the exhibition of presence. One of these being the nature of the embodiment and the set of gestures required for proper evaluation of the overall presence of the remote operator and how to ensure that the proper context is created, by also providing audio support to the depicted gestures. Furthermore, some elements with regards to the design of an experiment for evaluating general presence via robot avatars were also found, one of which being the set task that the participants are going to perform and the method of analysis of the response from the participants.

7.1.2 Hypotheses

As with the previous chapters, a number of sub-hypotheses based on the main ones described in Section 2.4.2, but with respect to social presence exhibited by tele-presence robotic avatars, have been devised. In the case of social presence, the level which an entity is able to exhibit, is usually with respect to a standard (Adalgeirsson & Breazeal, 2010). Using the criteria for the evaluation of social presence proposed by Biocca, Harms, and Gregg (2001), and their values for face-to-face conversation which was mentioned as being the 'gold standard' in social presence, a method of evaluation and a baseline for social presence was obtained. Also, since gesture reproduction is an important element of social presence, as shown by Adalgeirsson and Breazeal (2010), social presence for a robot avatar will be best evaluated when a robot that is able to portray gestures from the tele-operator is used.

Furthermore, in Chapter 4 and Chapter 5, it has been shown that the implemented control scheme used for the humanoid robot embodiment allowed it to provide measurable gains with respect to the non-humanoid robot using its normal control scheme. This together with the fact that the integrated control scheme, used to control the robots in this chapter is comprised of the robot pointing generation scheme a similar trend in the results is again expected. Therefore, when using the social presence evaluation criteria from Biocca et al., a significant difference is expected to be seen in the level of the social presence exhibited by the different robots (Biocca, Harms, & Gregg, 2001). This results in the following sub-hypotheses:

H 7.1 The level of social presence experienced will be significantly higher for the humanoid robot compared to the non-humanoid robot.

H 7.2 The social presence exhibited using a humanoid tele-operated robot, with an integrated control scheme, would make the presence comparable to face-to-face interaction.

Consequently, while taking into account the work of Adalgeirsson and Breazeal (2010) with special focus on the expanded set of factors rather than just the social presence factors, from Biocca et al. (2001), a similar trend is expected. These factors would allow a more comprehensive measure of the interaction by evaluating values for a set of dependent variables. Consequently while these dependent measures include the social presence, they also expand to add some metrics that can be used to evaluate the level of cognitive presence as well. Since a similar task to the one used in the work of Adalgeirsson and Breazeal (2010) is going to be used, an evaluation of the two robot conditions using the dependent variables can also be used to evaluate their overall presence. Also, considering that the measures for social presence are already being evaluated as part of the previous sub-hypotheses, the following sub-hypotheses were formulated:

H 7.3 The participants will trust the partner more when he is being embodied by a humanoid robot than when a non-humanoid robot is used.

H 7.4 The level of engagement within the group will be higher when a humanoid robot is used as a local embodiment compared to when a non-humanoid robot is used.

H 7.5 The cooperation between the group members will be better when the remote member is embodied by a humanoid robot than a non-humanoid one.

H 7.6 The participants will find the experience more enjoyable when a humanoid robot is used to portray the actions of the remote member when compared to a non-humanoid robot.

H 7.7 The use of a humanoid robot embodiment for the remote group member will allow for a higher level of persuasion when compared to a non-humanoid robot being used.

7.2 Method

In this chapter, the focus is on the level of presence exhibited by robotic embodiments or avatars, portraying the gestures of a remote participant, in a distributed group environment therefore, both a proper scheme and an appropriate task were required. The scheme, that is to be used, is meant to control the avatars and enable the gestures from the remote member to be appropriately displayed to the local one using robot avatars. This scheme builds upon the work carried out in the previous chapters where individual schemes aimed at portraying gestures of remote participants via robot avatars were described and evaluated. On the subject of the task that is going to be used in the experiment, an appropriate task and a method of evaluating the results were found while perusing works on social presence. The task makes use of the desert survival task as proposed by Lafferty, Eady, and Pond (1970) where the participants have to agree on how to prioritise items for survival in a desert. Using this scheme and this task, the experiment could be carried out. However, prior to running the main study, a pilot study was first conducted to verify that the implementation of both the control scheme and the task were successful.

7.2.1 Participants

Four participants took part in the pilot study (2 females, 2 males); aged 24 - 29 (M = 26.3, SD = 2.22).

For the main study, a total of 18 participants took part in the experiment (2 females, 16 males); aged 20 – 44 (M = 28.4, SD = 5.72). The 18 participants did not include any of the participants who took part in the pilot study to prevent their prior experience of the presence of the remotely located confederate via the embodiments to affect the collected data.

Due to the nature of the task that is being performed and the data that is being collected a further in depth analysis of the participants was performed based on their personality types while their level of knowledge of computer and robotics was used to provide background information on the distribution.

7.2.2 Material

To test the hypotheses stated in section 7.1.2, a suitable system was required. In Figure 7.2-A, a top-view of the implemented system is shown. As with Figure 5.2-A, the remote user in the top left corner of the figure is a remote user and is not physically present in the same environment and the actions of the remote user are being portrayed by a robot.

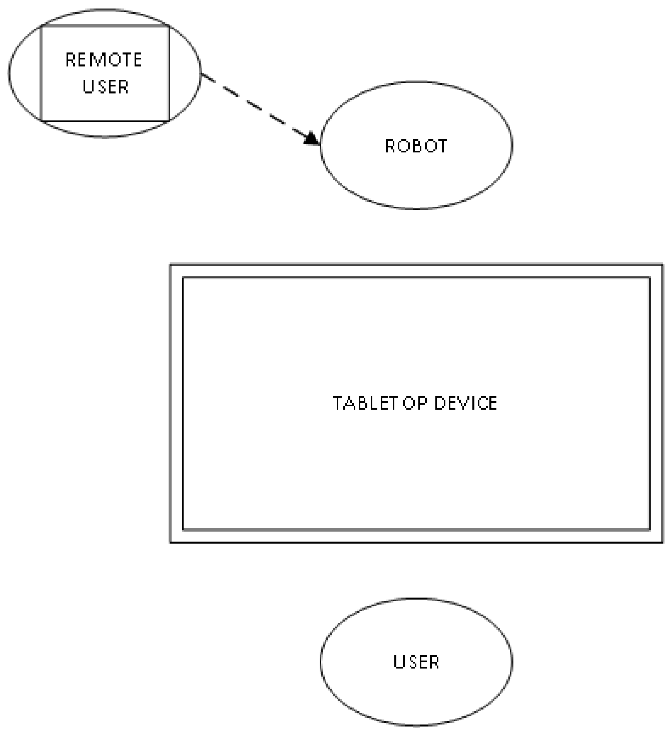

Figure 7.2-A: System Setup

Since the implemented system is to be used to evaluate the level of presence that different types of robots, each portraying gestures in accordance with an integrated scheme, some key components were required. These included, two tabletop interactive devices (TIDs), on which the task is displayed, different types of robots and the components of the capture system (Kinect, Data

Glove and Patriot Sensor), previously described in section 6.2.1. The two TIDs were comprised of a real one, the Microsoft Surface (See section 3.3.1) and a virtual one implemented using the Horizontal Screen (See section 3.3.2) and the Kinect described in section 6.2.2.1. Both of the robots used, the humanoid RoboThespian (See section 3.2.1.1) and the non-humanoid Cyton arm (See section 3.2.2.2), along with their control modules, were tasked with portraying the actions, similarly to Chapter 5, and also the gestures of the remote user. Audio support devices, similar to section 5.2.2.1, were placed at each location to allow audio communication, using Skype (See section 5.2.2.2), between the two locations. As in Chapter 5, a video recording device was also used to capture the experience of the participants.

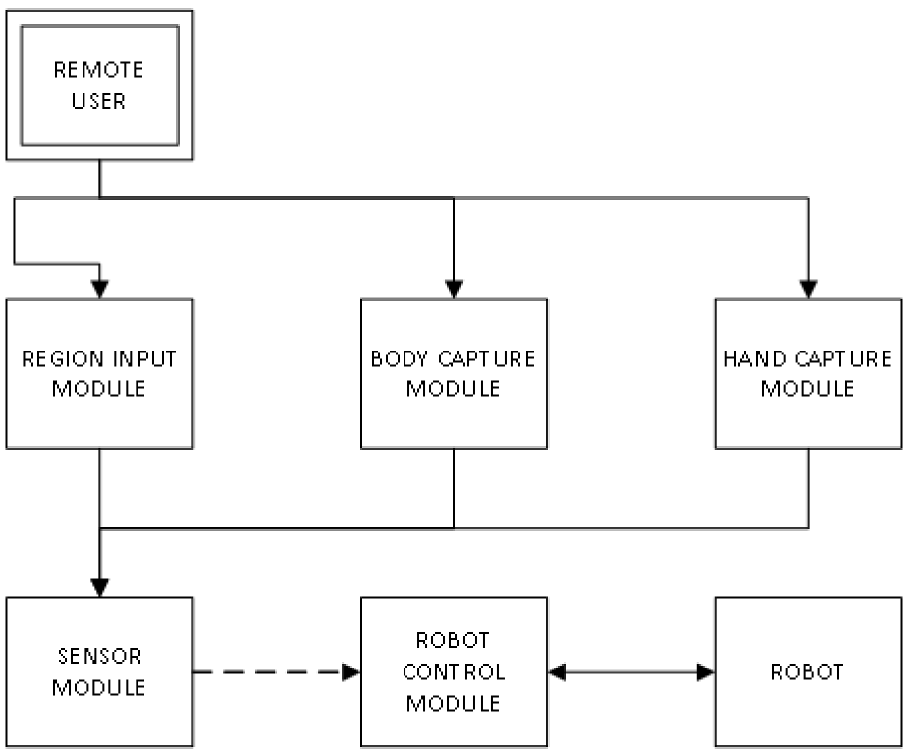

Figure 7.2-B: Capture and Output Components

In Figure 7.2-B, the main components of the system are shown. In this version of the system, there are three capture modules, (region input module; body capture module, for the pose of the body; and hand capture module, for the configuration of the hand), with two of them (body capture module and hand capture module) having been previously described in section 6.2.1. The final module, region input module, was implemented using the Kinect (See section 6.2.2.1), while the hand capture module was implemented using both the Data Glove (See section 6.2.2.2) and the Patriot orientation sensor (See section 6.2.2.3).

7.2.2.1 Capture Integration Scheme

In Figure 7.2-B, the capture system is portrayed with respect to its main components. The body capture and hand capture modules are used to implement a gesture modification scheme as described in Chapter 6. One previously identified limitation of the scheme is that it is inadequate for preserving the target of deictic gestures. That is, while the use of the gesture modification scheme can successfully depict a pointing gesture produced on the robot, the object the robot ends up pointing to may not be the same object that the remote participant is indicating.

Taking into account that the pointing gestures were inaccurate, though successfully identified using this scheme and that the other gestures were also successfully identified, one method of dealing with this lacking is to replace the pointing gestures by another scheme. Earlier in this thesis, in section 4.2.2.5, a pointing scheme was specifically implemented that allowed a humanoid robot to successfully convey pointing gestures. Using this pointing scheme, together with the gesture modification an integrated scheme is proposed. As part of this proposed scheme, a specific module was necessary to capture the gesture from the individual. This capture module can be implemented in two ways. Firstly, the pointing gestures could be identified, using a gesture classification system, and there could be a switch between the individual schemes when the performer or remote user is actually performing a pointing gesture. Implementing this option would increase the latency of the system, due to the use of the classification system. The second option makes use of virtual boxing of the environment, that is, the segregation on a real environment into a number of virtual three-dimensional zones. These zones can then act as virtual switches when an object is inside or outside of them, this entails identifying when objects are inside of these regions. Another element that made this scheme necessary is the absence of finger positions in the skeletal tracking scheme. The skeletal tracking scheme only implements accurate arm joints until the wrist joint and information about the position of the parts of the hands, specifically the fingertips, is inaccessible from this scheme. This is an acceptable option here since the use of pointing gestures is with respect to objects present on TIDs. This option was selected and implemented, using the Kinect as the input device, generating the region input module.

Unlike the skeleton stream, that is part of the Kinect SDK and that was used for the body pose capture module, the Kinect does not possess a region stream, so the region state identification was implemented using the depth image stream. As pointed out in section 6.2.2.1, the images obtained from this stream are made up of pixels, each representing the depth in a specific direction. Using the depth data from a depth image a representation of the environment can be produced. This representation can be enhanced to become a virtual duplicate of the current environment when taking into account both the orientation and position of the Kinect, with respect to the environment it is present in. Unlike the real environment that is continuous, this virtual representation is a ‘point cloud’ and it is made up of a series of individual points, each having an x, y and z coordinate, analogous to how a cloud is made up of individual droplets of water. This ‘point cloud’ adheres to the surfaces in the environment that are visible to the Kinect and creates a virtual representation of the environment. The positions of each of the individual points in the cloud are obtained using Equation 7.2-i.

Equation 7.2-i: Real world coordinate of point

$$p_i = t_o\big(o_k, s(m_i, v_i)\big) + p_k$$

In Equation 7.2-i above, $\mathrm{p_i}$ is the corresponding real world coordinate of a pixel i; $\mathrm{p_k}$ is the position of the Kinect; $\mathrm{o_k}$ is the orientation of the Kinect; $\mathrm{v_i}$ is the depth value of the pixel in the image, $\mathrm{m_i}$ is the position of the pixel in the image matrix; s is a function that calculates the position of the point in the coordinate frame of the Kinect after applying the surface transformation, it takes in the position of pixel in the image matrix and the depth value of the pixel in the image; and $\mathrm{t_o}$ is a transform function that calculates the relative position in the real world coordinate frame of the point with respect to the Kinect and takes in the orientation information and the position of the point in the coordinate system of the Kinect.

For an example of the Kinect reconstructing the environment see Appendix B.2.

From these point sets and a number of virtual three dimensional zones, it is possible to know whether an object in the environment is inside a specific region. Using this method two virtual regions were created above the physical horizontal screen. The positions, dimensions and orientations of the regions were calculated in order for them to adhere to be with respect to the horizontal screen. The first region referred to as the ***Touch*** region was a layer directly above the horizontal screen and was about 2 cm thick. The second region referred to as the ***Hover*** region was a layer directly above the Touch region and was about 3 cm thick. In Figure 7.2-C, a sideways depiction of the regions is provided together with the labelled regions and the positions of the main components.

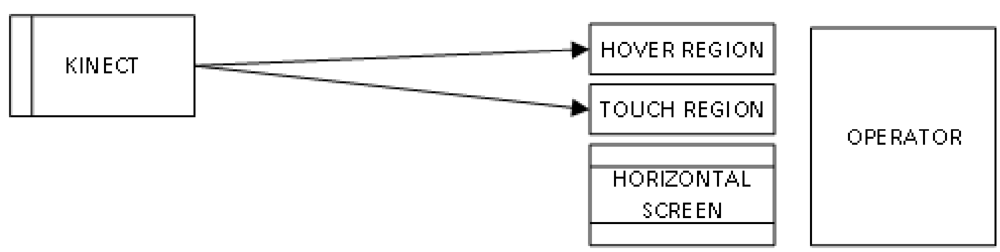

Figure 7.2-C: Side View of the Interactive Regions

While these regions could be simply used as triggers for when an object is inside of them, a more elaborate implementation was used. An analogy with the use of these regions can be drawn using mouse cursors and button controls from Graphical User Interface (GUI) designs. Starting with the hover region, this region can be used as a controller for a mouse cursor where by the user can use it to point to area of interest on the screen. Consequently, and using a parallel between the use of the horizontal screen and the client area of a form, in the hover region the section of the screen where the finger is present is as important as the area the cursor could designate without actively interacting with form elements. So, for the hover region in particular, finding the correct coordinate of the fingertip was important. As for the touch region, this region acts similarly to a mouse click on a button control, whereby it can trigger an interface event, thus allowing the user to select specific objects on the workspace on the screen.

In this implementation, these regions also have a more practical purpose. Using the hover region in conjunction with the touch region provides two benefits. First of all, the hover region can be used as an early trigger element for starting the pointing on the robot avatar, since the robot has to switch between the schemes and move into position. Thus, the motion of the robot avatar can be smoother and allowing the participant to correctly gauge the area of interest that the remote individual is designating. Moreover, the hover region acts as a safeguard to the touch region. Since only areas of the touch region will be activated by corresponding areas in the hover regions possible artefacts that will only be present in the touch region can be ignored, allowing for only pointing gestures to be part of the interaction.

The implementation of the area of interest control in the regions involved an averaging of the positions of points that were inside the region in order to find where the intruding object was with respect to the region, that is, the relative position of the intruding object with respect to the coordinate frame of the region. Using the above scheme, with the Touch region, the remote person would be able to virtually touch an object by moving his/her finger in the space inside the region while the coordinates of the contact point would be available from the coordinate frame of the region. Consequently, the contact point obtained from the region would have an x, y and z coordinate with respect to the region, that is each component of the coordinate would be ranged between -0.5 and 0.5. Using the x coordinate as measured from the sideways axis, the y coordinate as the depth axis between the individual and the sensor, and the z coordinate as the vertical height above the screen, the position of it with respect to any form of GUI control can be easily obtained while ignoring the z component. This allows for the pointing to be accurately captured and translated into the coordinate system of the robot avatar and TID in the local environment, thus ensuring that the robot avatar is designating the correct object.

7.2.2.2 Using the Cyton Arm

The hardware setup, in the local environment of the TID user, in this chapter is quite similar to the one used in Chapter 5 with the major difference between the two hardware setups being the non-humanoid robot used. In Chapter 5, the ABB arm was used (See section 3.2.2.1) while in this chapter the Cyton arm is used instead (See section 3.2.2.2). There are two main reasons why the ABB arm was replaced with the Cyton arm.

First of all, the function of the non-humanoid robot between the two experiments is different. The function of the non-humanoid robot, in Chapter 5, was limited to pointing to specific areas of the TID whereas in this chapter, the non-humanoid robot needs to portray a much wider range of gestures. The fact that the Cyton arm has more degrees-of-freedom and also possesses a gripper makes it a more practical option than the ABB arm, for the task at hand. Furthermore, the scheme developed in Chapter 6 to allow gestures captured from a remote participant to be reproduced at a different scale would be more suitably implemented on the Cyton arm.

Secondly, there is a safety component that needs to be taken into consideration. In Chapter 5, the ABB arm was used to point in a specific manner to different points on the TID, while this was carried out rapidly and accurately, the types of movements that were carried out by the robot were finite and pre-determined making safety less of an issue. The use of the robots in this experiment, being to portray a series of gestures from the remote participant and do so as quickly and accurately as possible makes safety a major concern. Since the Cyton arm moves at a low average speed and does not have much mass, it is a safe option for carrying out the gestures being performed by the remote operator, it was therefore selected as the non-humanoid robot embodiment for this part of the investigation.

7.2.2.3 Controlling the Cyton Arm

A control module was implemented in order to control the Cyton arm. This module uses a version of the integrated control scheme similar to the one used to control the movements of the humanoid robot. This integrated control scheme is made up of the pointing scheme from Chapter 4 and a modified version of the gesture modification scheme from Chapter 6.

Since the gesture modification scheme was also implemented on this robot in order for it to be able to portray gestures from the remote operator, segments of this robot were specifically selected to represent parts from the operator. Due to the form of the robot, only the movements of one arm could be represented by the robot. Nevertheless, the segments of the robot were used to represent the component of the torso and the right arm of the operator. The parts of the operator that were portrayed by the robot included the torso and hip, the right collarbone, the right upper arm, the right lower arm and the right hand. These segments were comprised of the links and degrees of freedom. These joints are depicted in Figure 7.2-D with a side comparison with the corresponding joints on a human being.

With these segments in place, the gesture modification scheme described in the previous chapter could be used to control the robot and allow it to portray the gesture of the remote operator. Using this configuration, most of the degrees of freedom of the human arm are supported by the robot except for the humeral rotation and the wrist movements while radial rotation is still supported. This lack of support for the humeral rotation, similar to the humanoid robot, was mitigated by the use of the joint in the hip to allow the robot to point in the right direction. It is to be noted that one important difference between the two versions of the gesture modification scheme is with respect to the use of the fingers. Since the Cyton arm is limited to a two-finger gripper, an average of the state of closure of the fingers was used as a control input for the gripper.

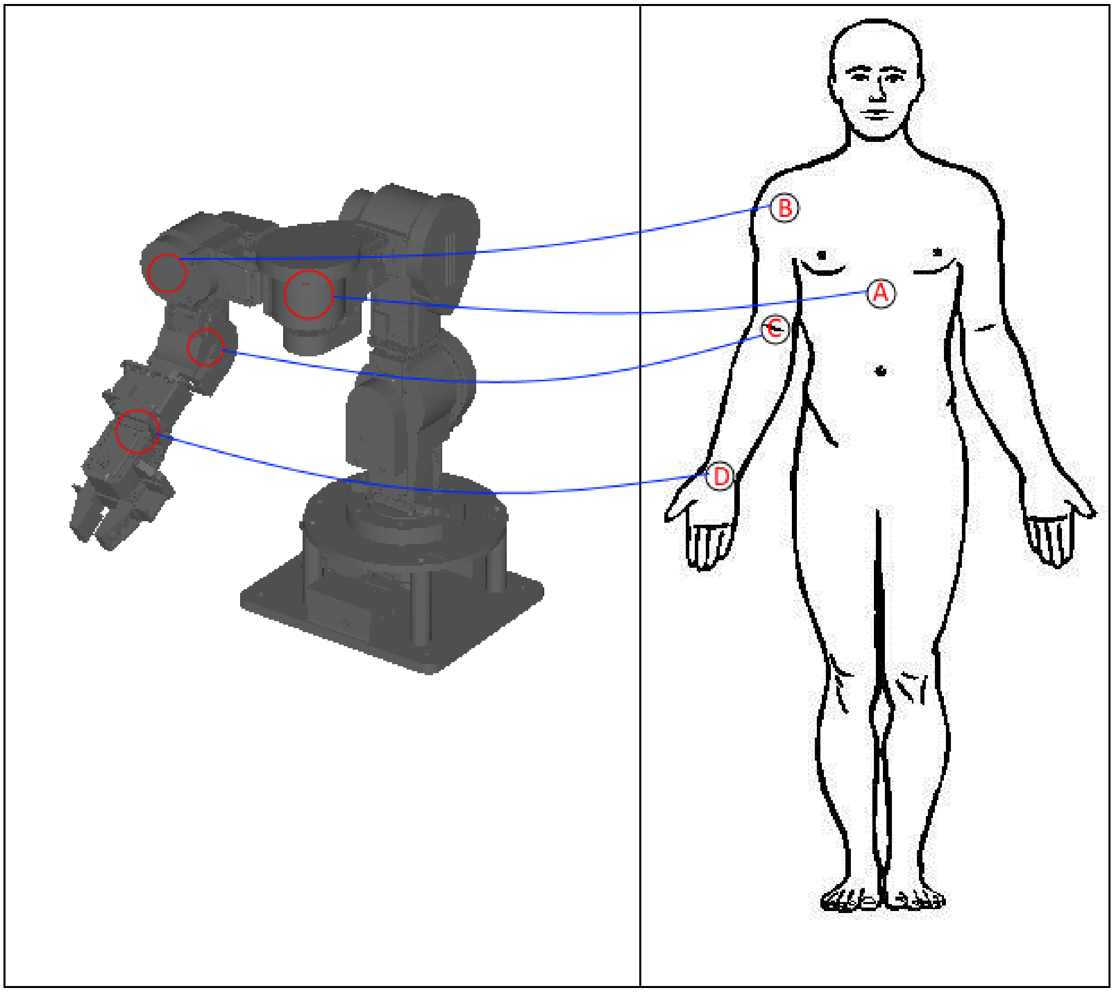

Figure 7.2-D: Joints of Cyton Arm and Human

Legend	
Joint A	Torso/Hip Rotation
Joint B	Shoulder Abduction/Adduction
Joint C	Elbow Flexion/Extension
Joint D	Wrist Pronation/Supination (Rotation)

7.2.3 Procedure

7.2.3.1 Pilot Study

In this chapter, a pilot study was carried out prior to the main study to verify that the implementation of the task using the current setup, in Figure 7.2-A, was successful, that the gestures being portrayed were able to be understood by the participants, and also to identify specific behavioural metrics for the analysis. As with the main study, a similar experimental procedure was used (See section 7.2.3.4), it involved distributed dyads, groups of two people, of participants collaborating over mirrored TIDs, with different types of robots being used as representation of the remote confederate and portraying his actions. This study was carried out four times with the participants experiencing different starting order for the different robot embodiments.

As with the main study, both the participants' interactions and the task state was captured. The former was captured using a video camera placed next to the participant to capture the participant and the robots within the frame. For the latter, screen capture software was used to record the state of the task together with the input from the different participants.

The result of this study confirmed that the current setup was successful but for the actual task implementation a few required changes were identified. One of these changes was the use of a predetermined script for the remote participant to go through in order to provide a more unified interaction style in the main study. For the gestures being portrayed by the robot it was noted that they were able to be correctly identified by the participants and also, as expected, that they were occurring within the right time frame for them to be synchronised with speech. One important behavioural metric that was identified was the focus area or the area of attention of the participants during the task and, although this will be further elaborated in the following section, this was observed primarily as the participant focussing on the head and face of the robot.

7.2.3.2 Video Analysis

As in Chapter 5 (See section 5.2.3.2), the video data captured using the camera, for the preliminary study, was analysed. However, while the analysis in Chapter 5 was aimed towards finding new metrics for evaluating the level of cooperation, the analysis in this chapter is aimed towards further supplementing the data obtained from questionnaires. While questionnaires are useful data capture tools, they are limited in providing only subjective participant responses. Behavioural measures give a more objective measure, and may aid analysis of system performance. Therefore while the questionnaire, similar to the one used by Adalgeirsson and Breazeal (2010), described later in this chapter, was used to gather most of the data, the video analysis was used to create a scheme with which the video data from the main experiments could be evaluated against.

The scheme used for the video analysis, in this chapter, contained two tracks:

- Participant track
- Task track

The attributes for the participant track were:

- Attention: Here it was noted what the participant was looking at (the TID, the robot body and in the case of the humanoid robot the head of the robot).

The attributes for the task track were:

- State: Here it was noted what the state of the task was, including whether it was running.

The selection of these particular attributes for these tracks was in order to provide context to the data gathered using the questionnaires. In earlier chapters, it was observed that participants had a tendency to glance at the head of the robot when the robot was performing a pointing gesture. While a similar behaviour is expected in the case where the robot is to point to objects, looking at the head of the robot may have another cause in this chapter. Since the human face is used as a medium to express a wide variety of sentiments (Oberman, Winkielman, & Ramachandran, 2007), and the task may involve a degree of feeling exchanged between the participants, looking at the head of the robot may also be due to the assessment of the mood of the remote participant through the robot. Therefore, using it as an attribute for the video coding scheme of the participant track will allow for the analysis of such behaviour. In order to place the behaviour in the proper context it was important to know at what the stage of the interaction this behaviour was observed, consequently an attribute from this was included as part of the video coding scheme.

7.2.3.3 Task

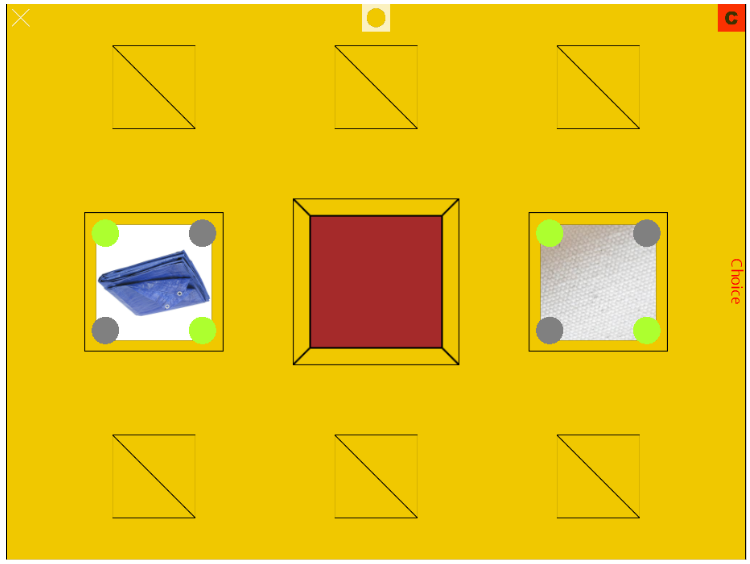

Figure 7.2-E: Desert Survival Task

As mentioned previously, the task that was used for the experiment was the desert survival task, proposed by Lafferty et al. (1970) and adapted by Takayama et al.(2009) in their study on influence from robots and also used by Adalgeirsson and Breazeal (2010), and Biocca et al. (2001), to investigate social presence (See Figure 7.2-E). The desert survival task is a group task where individuals have to work together in order to select from a series of items which ones are the most useful for surviving. While the purpose of it in the context of evaluating social presence in different settings ranging across media of interaction is the same, the structure of the task is known to vary between experiments.

There are several variations to the desert survival task (Adalgeirsson & Breazeal, 2010; Biocca, Harms, & Gregg, 2001; Lafferty, Eady, & Pond, 1970; Takayama, Groom, & Nass, 2009), the variation used here is the paired item approach used by Adalgeirsson and Breazeal (2010). This desert survival task was made up of 12 items grouped into 6 pairs (canvas/tarp, chocolate/water, mirror/compass, flashlight/matches, knife/gun, and map/book). These 12 items, arranged in 6 pairs, require 6 choices to be made, as to which of each pair is most useful for survival, with each item providing certain advantages and disadvantages. The 6 choices were arranged sequentially, that is, only one choice was available to the participants at any given time and the choices were accessed in the same sequence by each participant. The participants had to decide which item in a pair was most appropriate and select it by dragging it into the selection box. The remote operator would then agree with or contest the decision. This allowed for a discussion after which the item that was agreed upon was selected. The following choice was then presented while the interface depicted the selected option of the previous choices in order to decrease the cognitive load on the participants. The task was only completed when an option was selected for each of the 6 choice.

7.2.3.4 Experimental Procedure

For each of the sessions, the system was set up with the participant and active robot on opposite sides of the TID while the remote user was in a different location, with the capture system and the horizontal screen (See Figure 7.2-A).

During each of the sessions, the actions of the remote collaborator were portrayed by the active robot embodiment. Using a within subject design for the experiment allowed each participant to experience both embodiments, therefore the task was divided into two sections with each section consisting of three choices and involving a specific embodiment.

With the robot avatars tasked with portraying the gestures, an important part of the interaction used the deictic gesture implementation scheme and allowed for the selection of the object to be in line with the overall interaction. The deictic gestures were also used during the introduction phase of the interaction to designate the objects prior to the first selection of the object. In addition to the deictic gestures other conversational gestures, primarily and beat gestures were also used by the remote participant. The portrayal of the gestures from the remote participant was part of the script provided for the remote participant.

As mentioned earlier, in order to allow the experience to be more unified, an interactive script was created for the operator to follow, which included the arguments to use to respond to the choice the local participant made. This allowed for the factors impacting the interaction to be limited, as much as possible, to the robots and the gestures portrayed, while having the added advantage of decreasing the cognitive load of the operator.

To control for ordering effects pseudo-randomisation of the order of the robots and the choice to be contested by each was used, counter-balanced across participants. Hence, half of the participants experienced one robot order and half the other order, and each participant experienced a different combination of robot and choice contesting. In order to increase consistency between the different sessions, the same remote operator was used for each session of the experiment.

As with the pilot study, the experience of the local participant was captured using both a video camera present in the local area of the participant and screen capture software. A new addition developed from the pilot study was an automatic logging system that logged the game interactions and also provided some of the metrics data.

The data capture of the experiment also included a questionnaire (see Appendix A.3) that the participant had to fill in at different points throughout the sessions. This questionnaire included different sections spanning from the background information of the participant to how the participant felt about each embodiment. This questionnaire was provided to the participants prior to the start of the experiment along with instructions on both how to fill it and also how to perform the task; the instructions provided to the participants included both an instruction sheet and verbal instructions.

Finally, in one of the sections of the questionnaire, the session code was filled in; this session code included information on the ordering of the robot and which choices would be contested. This code, allowed the sessions to be double blinded, i.e. the presentation order and choice contesting was not used in the result analysis. The code was obtained from the randomisation scheme previously mentioned and generated using a random number generator and an exhaustive table.

7.3 Results

7.3.1 According to the Population Demographic

The participants used in this experiment had a relatively high knowledge of technology from the data ranked on a seven point scale. With their knowledge of computers (M = 5.83, SD = 1.65) being higher that their knowledge of Robotics (M = 5.33, SD = 2.30).

Apart from their knowledge of technology, another type of demographic information obtained from the questionnaire is personality type. This is going to be used later in the results section to analyse the impact that some personality traits have on the persuasion. However, in this section a brief description of the demographics is provided. Using the Big 5 personality test measured on a 50 point scale (Goldberg, 1999), it was found that the population was mostly Extraverted (M = 33.6, SD = 6.23), Agreeable (M = 38.1, SD = 5.34), Conscientious (M = 34.4, SD = 6.34), Emotionally Stable (M = 32.7, SD = 5.95), and Intellectual (M = 39.6, SD = 4.95).

7.3.2 According to the Social Presence Factors

The data obtained from the social presence sections of the questionnaire was first scored, and the scored data was analysed with respect to the factors put forward by Biocca et al., (2001). Using a similar data collection scheme as Biocca et al., the data was collected using a 7 point Likert scale and is represented with respect to the same scale. From the analysis of the data, Figure 7.3-A and was obtained.

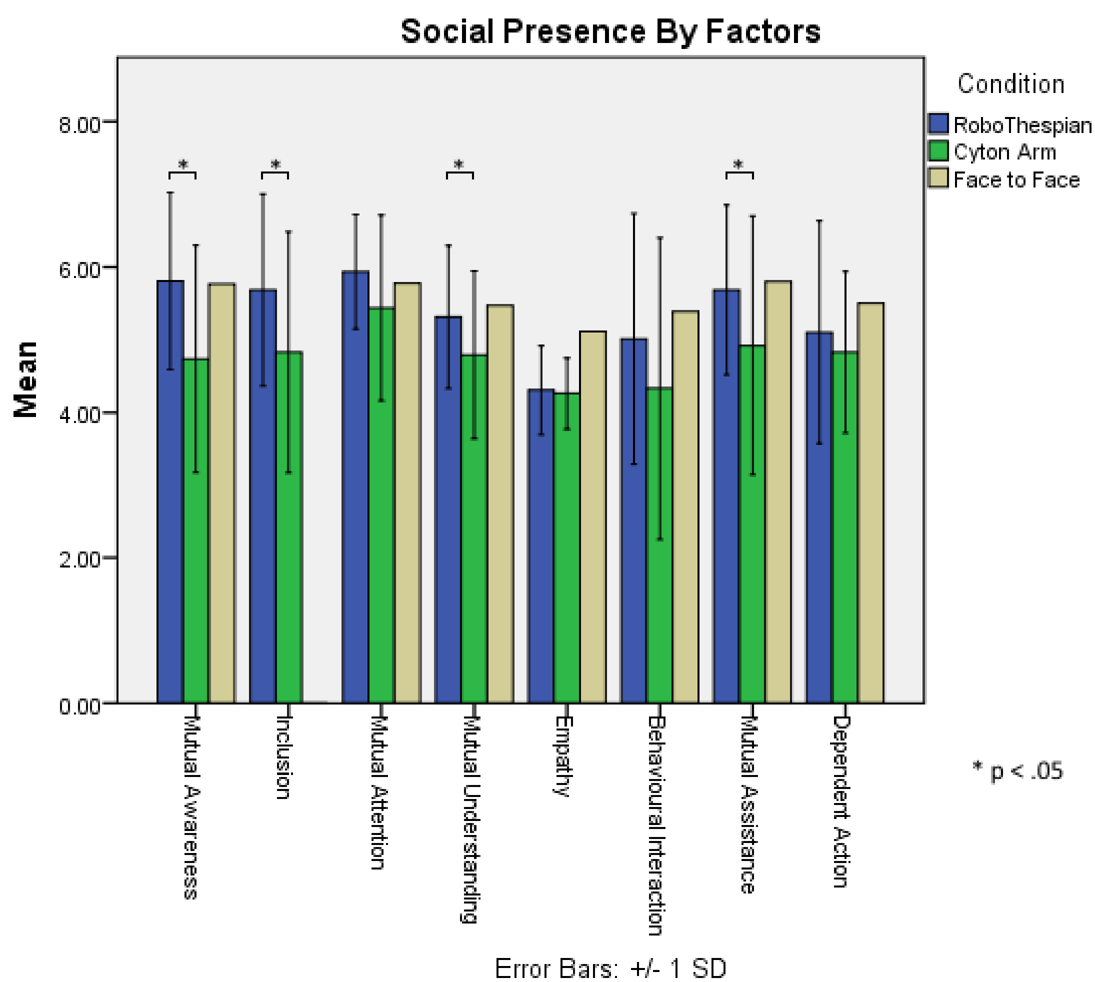

Figure 7.3-A: Social Presence by Factors

These representations also include, the Face-to-Face values for each of the factors, provided by Biocca et al., of the same activity for comparison (Biocca, Harms, & Gregg, 2001). It is to be noted that since no value was provided in Biocca et al. (2001), for the 'Inclusion' factor for Face-to-Face interaction, there is no corresponding representation in Figure 7.3-A.

Since all the data obtained is of the interval type and the same group was used to get data from both conditions, one possible test to use is the dependent T-Test. However, in case of data violating the assumption of Normality (tested using a Shapiro-Wilk's test of Normality (Shapiro & Wilk, 1965), a Wilcoxon signed-rank test was used instead. This procedure was performed for each of the factors and for the overall social presence as well. It is to be noted that for the Face to Face condition the data obtained and represented in Figure 7.3-A only contained mean values and did not have standard deviation data. This lack of standard deviation value prevents effective comparison between all three conditions, which would have involved repeated measures ANOVAs and Friedman tests instead of dependent T-Tests and Wilcoxon signed-rank tests respectively.

Starting with the Mutual Awareness factor, a dependent T-Test was used to analyse the data for each condition: RoboThespian ($M = 5.81$, $SD = 1.22$) and Cyton Arm ($M = 4.74$, $SD = 1.57$). The result of which determined that the values for the Cyton arm condition were found to be significantly different when compared to those of the RoboThespian condition ($t\ (17) = 2.39$, $p = .029$).

For the Inclusion factor, a dependent T-Test was used to analyse the data for each condition: RoboThespian ($M = 5.68$, $SD = 1.32$) and Cyton arm ($M = 4.83$, $SD = 1.66$). The result of which determined that the values for the RoboThespian condition were found to be significantly different when compared to those of the Cyton arm condition ($t\ (17) = 2.14$ $p = .048$).

For the Mutual Attention factor, a Wilcoxon signed-ranked test was used to analyse the data for each condition: RoboThespian ($M = 5.94$, $SD = 0.784$) and Cyton arm ($M = 5.44$, $SD = 1.28$). The result of which determined that the values for the different robot conditions were found to be not significantly different from one another ($Z = 1.78$ $p = .075$).

For the Mutual Understanding factor, a dependent T-Test was used to analyse the data for each condition: RoboThespian ($M = 5.31$, $SD = 0.986$) and Cyton arm ($M = 4.79$, $SD = 1.15$). The result of which determined that the values for the different embodiment conditions were found to be significantly different for each embodiment ($t\ (17) = 2.14$, $p = .047$).

For the Empathy factor, a dependent T-Test was used to analyse the data for each condition: RoboThespian ($M = 4.31$, $SD = 0.611$) and Cyton arm ($M = 4.26$, $SD = 0.490$). The result of which determined that the differences between the robotic embodiments were not significantly different ($Z = .42$, $p = .673$).

For the Behavioural Interaction factor, a dependent T-Test was used to analyse the data for each condition: RoboThespian ($M = 5.01$, $SD = 1.72$) and Cyton arm ($M = 4.33$, $SD = 2.08$). The result of which determined that the values for the RoboThespian were not significantly different from those of the Cyton arm condition ($t\ (17) = 2.01$, $p = .061$).

For the Mutual Assistance factor, a Wilcoxon signed-rank was used to analyse the data for each condition: RoboThespian ($M = 5.68$, $SD = 1.17$) and Cyton arm ($M = 4.92$, $SD = 1.78$). The result of which determined that the values for the Cyton arm were significantly different from those of the RoboThespian condition ($Z = 2.08$, $p = .038$).

Finally for the Dependent Action factor, a Wilcoxon signed-rank test was used to analyse the data for each condition: RoboThespian ($M = 5.10$, $SD = 1.53$) and Cyton arm ($M = 4.83$, $SD = 1.11$). The result of which determined that the values for the RoboThespian condition were not found to be significantly different from those of the Cyton arm condition ($Z = 1.13$, $p = .260$).

For the overall level of social presence a repeated measures ANOVA was used to analyse the data for each condition: RoboThespian (M = 5.31, SD = 0.565), Cyton arm (M = 4.76, SD = 0.394), and Face2Face (M = 5.55, SD = 0.254). The result determined that the mean value for social presence differed statistically significantly between the different conditions ($F(2, 12) = 24.8$, $p < .001$). Post-hoc pairwise comparisons with dependent t-tests were conducted, with Bonferroni correction, the results of which showed that the use of different types of embodiments caused a difference on the level of social presence. While there was a decrease in the overall level of social presence between the Face2Face and the RoboThespian conditions, the difference was found to be not statistically significant ($p = .301$). However, this is the only case where the difference is not found to be significant as when compared to the Cyton arm embodiment both other conditions were found significantly different. First when comparing the Cyton arm with the RoboThespian condition, the level of social presence was found to be much lower and with a notable significance in the difference ($p = .014$). Finally, when comparing the Cyton arm with the Face2Face condition, the difference was found to be even bigger than the previous condition and the significance to be even more pronounced ($p < .001$). It should be noted that despite these results obtained any conclusion drawn from them with respect to the Face to face condition could be conclusive since the values used for this condition were obtained from the work of by Biocca et al., (2001) rather than found as part of this study.

7.3.3 <u>According to the Dependent Variables</u>

As mentioned previously, apart from social presence, two other types of presence can also be exhibited in tele-presence settings. In the work of Adalgeirsson and Breazeal, a set of measures were used to assess the usefulness of the tele-presence experience; the measures included the social presence measures, from Biocca et al. (2001), but also other measures that took the cognitive aspect of presence into account (Adalgeirsson & Breazeal, 2010). These non-social presence measures were also included as part of the questionnaires that were given to the participants.

In the model for the dependent variables proposed by Adalgeirsson and Breazeal (2010), a total of eight components were selected, the first three components are linked to social presence and act as a summary for the measurement of level of social presence. The first three dependent measures therefore reuse the eight components from the work of Biocca et al. (2001), with each measure including one or more of the eight components of the previous section. As mentioned previously, the individual questions for these measures were collected using a 7 point Likert scale.

Apart from the social measures, most of the other measures were also captured using rating scales apart from the last measure which was obtained directly from the logging system and was captured using a ratio scale. The rating scales used for these measures used 8 point scales and the fourth measure similar to the first three measures were also set using a number of individual components. This measure, the trust measure, is based on the work of Wheeless and Grotz (1977) and is implemented using the Individualised Trust Scale (ITS) and a set of 15 factors.

Since the first three measures are based on the social presence factors from the work of Biocca et al. (2001), which have already been analysed in the previous section, the RoboThespian has a higher value for each of them. A summary of the results for each measure of social presence for each robot is depicted in Figure 7.3-B.

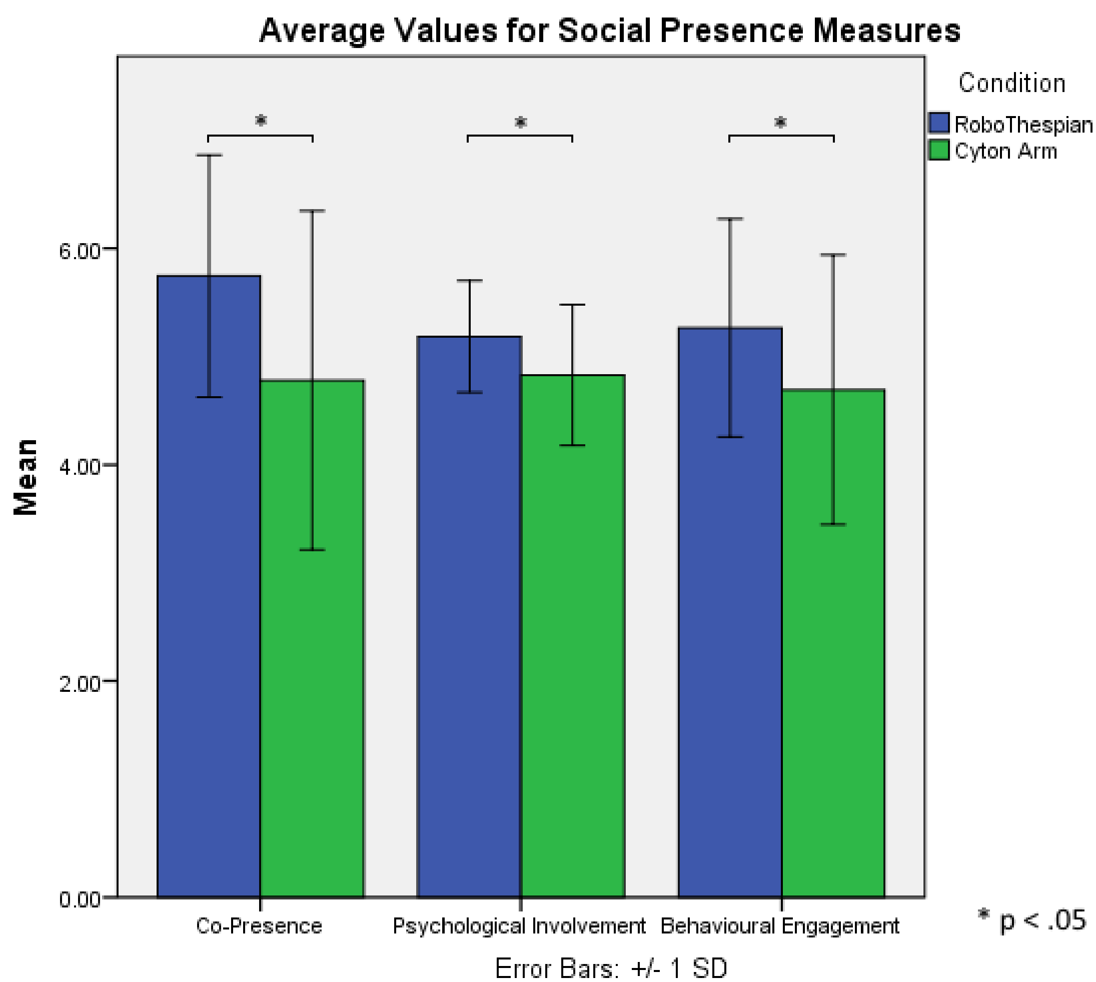

Figure 7.3-B: Average Values for Social Presence

Although the RoboThespian is expected to be significantly different due to the findings of the previous section, nevertheless the individual computed summaries were still tested for difference. A dependent (paired-sample) T-Test would be a suitable test to be used here due to there being only two conditions and the same participants having been used in each condition. Since the variables are measured on a continuous scale and the pairs are related, only normality and no-outliers need to be satisfied. However, in case of violation, a Wilcoxon Signed Rank Test, would be used instead. The factors were individually tested.

For the Co-Presence variable, a dependent T-Test was conducted comparing the conditions: RoboThespian ($M = 5.74$, $SD = 1.12$), Cyton arm ($M = 4.78$, $SD = 1.57$). The results showed that there was a significant difference between the two conditions, $t\ (17) = 2.20$, $p = .042$.

For the Psychological Involvement variable a dependent T-Test was again conducted comparing the conditions: RoboThespian ($M = 5.18$, $SD = 0.515$), Cyton arm ($M = 4.83$, $SD = 0.652$). The results of which showed that there was a significant difference between the two conditions, $t\ (17) = 2.81$, $p = .012$.

For the Behavioural Engagement variable, a dependent T-Test was conducted comparing the conditions: RoboThespian ($M = 5.27$, $SD = 1.01$), Cyton arm ($M = 4.69$, $SD = 1.25$). The results showed that there was a significant difference between the two conditions, $t\ (17) = 2.29$, $p = .035$.

Similarly to the first three measures used for social presence, the fourth to seventh measures were also analysed. The results of the analysis are shown in Figure 7.3-C.

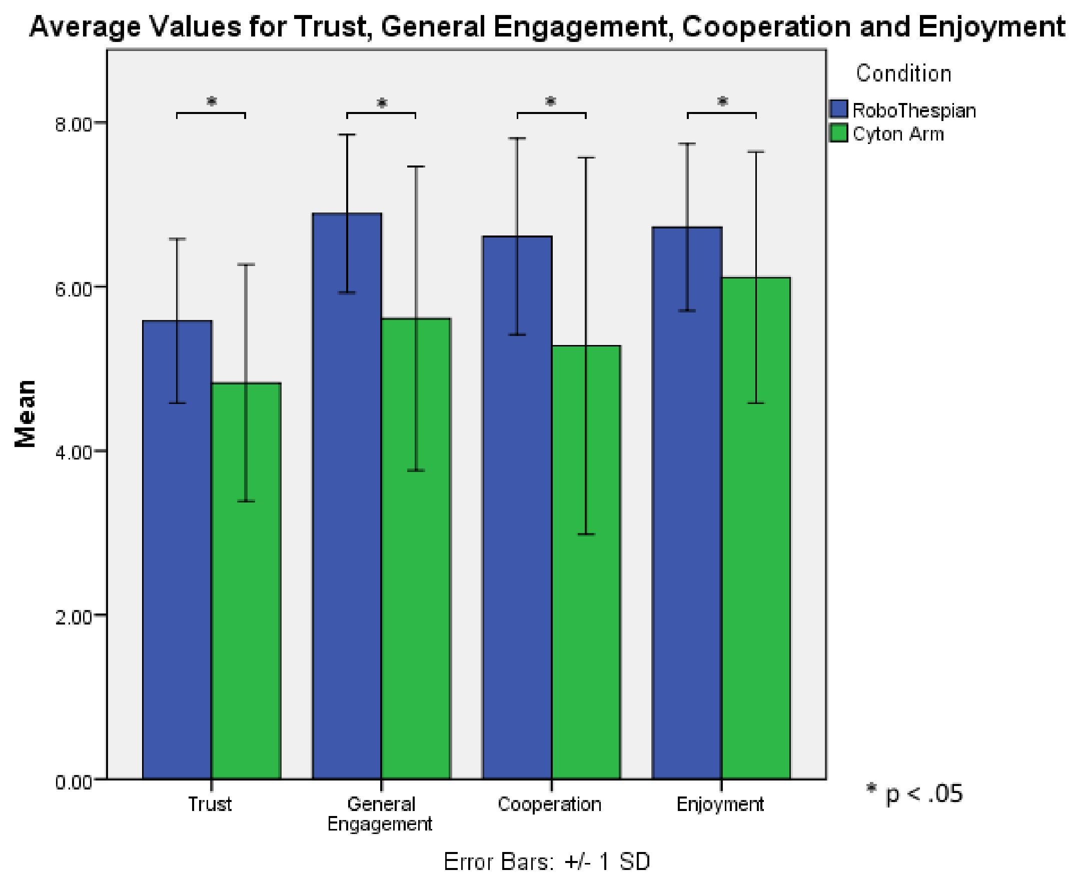

Figure 7.3-C: Average Values for Trust, General Engagement, Cooperation and Enjoyment

A dependent (paired-sample) T-Test was again selected to analyse the data for the different measures, across the conditions, and similar to the previous analysis, a Wilcoxon Signed Rank Test, would be used in case of violation of the assumptions.

For the Trust measure, after the values obtained were scored to obtain a ratio scale, a dependent T-Test was conducted comparing the conditions: RoboThespian ($M = 5.58$, $SD = 1.00$), Cyton arm ($M = 4.83$, $SD = 1.44$). The results showed that there was a significant difference between the two conditions, $t\ (17) = 2.94$, $p = .009$, with the RoboThespian showing a higher mean.

For the General Engagement measure, where the data was captured on a ratio scale, a Wilcoxon Signed-Rank test was conducted comparing the conditions: RoboThespian ($M = 6.89$, $SD = 0.96$), Cyton arm ($M = 5.61$, $SD = 1.85$). The results showed that there was a significant difference between the two conditions, $Z = 2.62$, $p = .009$, with the RoboThespian showing a higher mean.

For the Cooperation measure, a Wilcoxon Signed-Rank test was again conducted comparing the conditions: RoboThespian ($M = 6.61$, $SD = 1.20$), Cyton arm ($M = 5.28$, $SD = 2.30$). The results showed that there was a significant difference between the two conditions, $Z = 3.13$, $p = .002$, with the RoboThespian showing a higher mean.

For the Enjoyment measure, a Wilcoxon Signed-Rank test was again used comparing the conditions: RoboThespian ($M = 6.72$, $SD = 1.02$), Cyton arm ($M = 6.11$, $SD = 1.53$). The results showed that there was a significant difference between the two conditions, $Z = 2.33$, $p = .02$, with the RoboThespian showing a higher mean.

Finally, the eighth measure, which was obtained automatically from the logging system, was also analysed. The results of the analysis are shown in Figure 7.3-D.

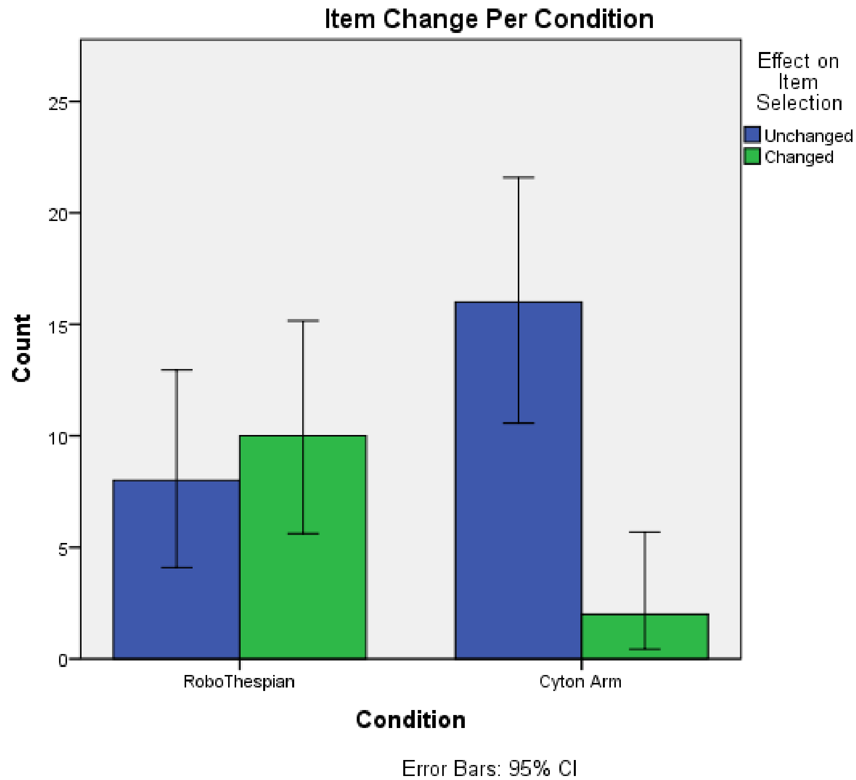

Figure 7.3-D: Item Change Per Condition

The last measure in this list is Item Change, it is a measure of the change in the item selected after the confederate argues against it. It is also a measure of persuasion from the robot embodied confederate. As previously mentioned, for each run of the experiment only one item was chosen to be argued for each robot. Hence, for this measure, the data is binomial, i.e., discrete and dichotomous. Consequently, a McNemar's test was selected to analyse the data, the exact variant of the test is used due to the sample size.

The results of the Exact McNemar's test on the conditions: RoboThespian, and Cyton arm, showed that there was a strong statistical significant relationship, $p = .021$, between the type of robot used and the effect on item change, with the RoboThespian embodiment condition having a higher effect on persuasion.

A further analysis was carried out on this measure with respect to the personality of the different participants. For this analysis the specific traits from the Big Five were focussed on as being contributing factors to the Item Change. These traits are the Extraverted, Agreeableness, Conscientious, Emotional Stability, and Intellectual. The participants were categorised as Extrovert / Introvert, Friendly / Hostile, Careful / Careless, Stable / Unstable, and Intelligent / Unintelligent for each trait respectively. Using these category pairs, the personality trait data is discretised and becomes nominal thus making it suitable for further testing.

However, from a brief analysis of the data, for two of the traits, Agreeableness and Intellectual, the negative categories accounted for only one data point and therefore made up less than 6% of the population of participants. So for these two traits no further analysis was carried out since no significant result could be calculated. For the remaining three traits, Extraverted, Conscientious, and Emotional Stability, with more than 5 data points in each of the category for these traits, results could be inferred from the data for each trait.

Since the result of the above exact McNemar's test showed a significant relationship between the type of robot used and the effect on persuasion, follow up tests were performed, to verify whether the main trend is also valid for each category of the selected traits. Since each trait is composed of two categories, that are mutually exclusive, for each trait, the population was divided into two subgroups with each subgroup containing paired samples of the result on item change of the use of each robot. The follow up tests were each conducted using an exact McNemar's test since the samples are related and taken from a small population.

Starting with the first personality trait, the Extraverted trait, the results of the follow up test showed that the effect of the type of robot used on persuasion was different between the categories. For the extroverted participants, the relationship between the type of robot used and the effect on persuasion was strongly statistically significant ($p = .016$) but not in the case of introverted participants ($p = 1$), indicating that for the extroverted participants the type of robot used had an effect on the decision to change item.

For the Conscientious trait a similar trend was also found with the effect of the type of robot used on persuasion being different across the different categories. For the careful participants, the relationship between the type of robot used and the effect on persuasion was strongly statistically significant ($p = .016$) but not in the case of careless participants ($p = 1$). Similarly to the extraverted trait, a similar conclusion can also be drawn for the conscientious trait between the type of robot used and item change for the careful participants.

For the final personality trait, the Emotional Stability trait, a similar trend across categories was also observed. The results in this case showed that for the emotionally stable participants that the relationship between the type of robot used and the effect on persuasion was strongly statistically significant ($p = .031$) but that was not the case for the emotionally unstable participants ($p = .625$). This indicates that for the emotionally stable individuals, the type of robot has an impact on the decision to change the item.

Consequently, the personality of the participants as well as the type of embodiment contribute to the change in item. With participants who are either extroverted, careful or stable more likely to be persuaded to change the item than the antagonistic traits. However, for the agreeableness and intelligent traits, the results obtained were inconclusive. From the results obtained, it can be surmised that open and reflective individuals are more likely to be persuaded by a humanoid robot being used as a remote embodiment. It should be noted that while looking promising these results are tentative due to the limited size of the sample.

Finally, a Wilcoxon Signed-Rank test was used to analyse the values for the different measures with respect to the different conditions: RoboThespian ($M = 5.32$, $SD = 2.04$), Cyton arm ($M = 4.53$, $SD = 1.85$). The results showed that there was a significant difference between the two conditions, $Z = 2.52$, $p = .012$, with the RoboThespian showing a higher mean.

The above results strongly support hypotheses H 7.3, H 7.4, H 7.5, H 7.6, and H 7.7.

7.3.4 According to the Video Annotation

As mentioned previously, the interaction of the local participant with the tabletop interactive device and robots were also captured via a video camera. However, while such captured data, in previous chapters, were used as the basis to find the difference between conditions, in this chapter, the purpose of it is primarily to determine the behavioural metrics that is used for social presence. In addition to this, the captured data is also used to provide further support to the differences found using the other metrics.

The video coding scheme, developed following the preliminary study was applied to recordings of the participant performing the task in front of each of the robots. The results of the coding scheme were then qualitatively analysed to provide the supporting arguments for the main findings. A lot of focus was placed on the area of attention of the participant throughout the interaction and it was found that the participants shifted attention to the head of the RoboThespian during the interaction.

Prior to analysing the video for the behavioural metrics, the context was needed, this required analysis of the video in terms of time spent in seconds for each condition. A dependent (paired-sample) T-Test would be a suitable test to be used here with a Wilcoxon Signed Rank Test, in case of normality violation. A dependent T-Test was conducted comparing the conditions: RoboThespian ($M = 180$, $SD = 33.9$), Cyton arm ($M = 177$, $SD = 47.5$). The results showed that there was no significant difference between the two conditions, $t(17) = 0.159$, $p = .876$. Having established that there were no significant differences between the robot conditions, in terms of duration, the behavioural metrics can now be analysed within a proper context.

One encoding element that was part of the video coding scheme was the attention part, which was focussed on where the participant was looking at. Using the annotated videos, coded using the video coding scheme, both duration and times the participants looked at the robots can be obtained. Since there was no significant difference in the duration of the videos with respect to the robot duration of looking at the robot is going to be used. A dependent (paired-sample) T-Test would again be a suitable test to be used here while a Wilcoxon Signed Rank Test will be used instead, in case of normality violation. A dependent T-Test was conducted comparing the conditions: RoboThespian ($M = 113$, $SD = 21.0$), Cyton arm ($M = 51.9$, $SD = 12.8$). The results showed that there was a significant difference between the two conditions, $t(17) = 12.55$, $p < .001$. That is the participants spent more time looking at the RoboThespian than the Cyton arm during the experiment.

Moreover, it should also be noted that the participants also shifted attention by glancing at the face of the RoboThespian at the start of conversations with the remote participant and since a similar behaviour was not seen in the case of the Cyton arm, this unique behaviour for the RoboThespian was not able to be compared quantitatively between the two robot conditions.

7.4 Discussion

From a brief observation of the results obtained, displayed in Figure 7.3-A and Figure 7.3-B, it can be seen that, with respect to the level of social presence from the different embodiments, two main things seem to be apparent. Firstly, that the humanoid robot used as an avatar for a remote member exhibited a higher level of social presence than the non-humanoid robot was used. Secondly, that the values obtained for the humanoid robot were generally similar to the Face-to-Face condition while the non-humanoid robot displayed values that were markedly different. This would tend to lend some credence to hypothesis H 7.1 and also part of hypothesis H 7.2.

According to hypothesis H 7.2, the use of tele-operated robots, using the integrated control scheme, would make the level of social presence comparable to Face-to-Face interaction. This would be represented by an absence of significant difference between the values of the robot embodiments and the Face-to Face conditions for the different factors. From the results obtained it was found that despite the difference in the values for the Empathy factor, the RoboThespian condition was found to not be statistically different from the Face-to-Face condition, thus partially supporting the hypothesis H 7.2 at least where the RoboThespian is concerned. However, the difference between the values obtained for the Cyton arm condition and the Face-to-Face condition were found to be statistically significant thus not supporting hypothesis H 7.2, in the case of the non-humanoid robot. Therefore from the results obtained it would seem that hypothesis H 7.2 is only partially supported. Moreover, as mentioned earlier, the values for the Face-to-Face condition were obtained from the literature rather than from the experiment and therefore the comparison can only be used with caution since it is possible to have differences between the two set-ups and scripts.

Furthermore, according to hypothesis H 7.1, a higher level of social presence would be exhibited by the humanoid robot, RoboThespian, than the non-humanoid robot, Cyton arm. The results of the analysis support the hypothesis since a higher average value was obtained for the

RoboThespian condition than for the Cyton arm condition and also since there was a significant difference between the two robot conditions.

This is further supported by a statistical analysis of the data, for each of the three dimensions of social presence put forward by Biocca et al. (2001), where significant differences were found between the two embodiments. While this helps support hypothesis H 7.1, a comparison between the two embodiments, for the eight constituting factors, showed that this difference, between the two embodiments, was least expressed for the Empathy factor.

According to Ickes (1997), empathy can be defined as the ability of an individual to perceive and understand the thoughts and emotions of another entity. Moreover, empathy can sometimes also be defined similarly to sympathy since the person understanding the state of the other entity can also share the state without being in the same situation (Decety & Jackson, 2004). Having defined what empathy is, it is now important to find out why the level of empathy was comparatively low for both robot conditions and also why no significant difference was found between the two robot conditions in the case of empathy.

As mentioned previously, one of the premise for empathy is the ability to perceive the thought or emotional state of the other entity. While this perception can be done through different channels, including both audio and visual, one of the most basic medium is the use of the facial expressions (Trevarthen & Aitken, 2001). Even prior to the understanding of audio communication, facial expression can be used to transmit the state that an entity wants to convey (Trevarthen & Aitken, 2001). Since empathy is best transferred with a human face (Cole, 2001), research has been carried out into designing robots that can emulate facial empathic features in order to facilitate interaction between robots and humans (Kim, Kwak, & Kwak, 2009).

While the humanoid robot possessed a face, the face had a very limited set of emotional expressions and they were not used during the experiment, that is, the robot presented a neutral face. The non-humanoid robot did not have a face or even a head. The lack of facial expressions from the robot could be the reason why there was a limited perception of empathy by the participants. This would account for the lower level of empathy, since facial expression is a feature that is present in Face-to-Face interactions but was absent in both of the robot mediated conditions. Analysis of the video data from the experiment showed that the participants looked at the head of the robot from time to time during the experiment. While it can be argued that such behaviour is an indicator of the participant attempting to find more empathy from the robot, i.e. looking for facial expression that is not present, it should be noted that the head has another function. As mentioned previously, the head of the humanoid robot is used in the object indication process, where the head was used to convey information about the area of interest being designated by the robot, i.e. using gaze. Therefore, further investigation will be required to find out the exact impact of possessing a dynamic face, able to express facial gestures, on robot mediated tele-presence and also regarding the most effective way of conveying empathic information for robot mediated tele-presence. Moreover, a more emotionally involved task may be required since while reviewing the captured video data there did not seem to be a lot of emotional involvement from the participants. This lack of involvement could be either due to the nature of the task or the sample group since they tended to the analytical rather than friendly end of the agreeableness personality trait spectrum. Therefore in addition to the investigation into the need of facial gesture from the embodiment a more emotionally involved task with a more emotionally inclined individuals may be required.

Despite the difference in the values obtained for the social presence factors, for each robotic embodiment, being least pronounced for the empathy factor, its overall effect on the three social presence dimensions was negligible. These social presence dimensions were a summarised form of the eight factors, with each being computed from two or three individual factors; they were also included as part of the eight dependent measures. While social presence and its constituent dimensions are important elements of the dependent measures, these measures go beyond the social aspect of presence and attempt to evaluate the overall tele-presence experience. Therefore, while the three social dimensions, Co-presence, Psychological Involvement and Behavioural Engagement, were included, other measures such as Trust, General Engagement, Cooperation, Enjoyment and Items Change were also present.

The result of the analysis of the dependent measures, showed a significant difference between the values obtained for the two conditions, supporting hypotheses, H 7.3, H 7.4, H 7.5, H 7.6, and H 7.7, this being most noticeable for the Items Change measure. Considering that the same task was used for each condition then the contributing factor for the difference needs to be the robot-scheme combination. As mentioned previously, the control scheme used for the individual robots integrated two individual schemes, the gesture modification scheme that allows captured gestures to be implemented on the robots and the pointing scheme, for indicating areas of interest. While both robots used the same gesture modification scheme, simply adapted to their specific morphologies, each pointing scheme was individually based. While pointing was minimally used compared to the previous chapters, limited to two objects of opposite sides of the TID, it remains a contributing factor. Another factor that can have an impact on the dependent measures has to do with the perception of the robot; this factor incorporates both the form and size of the robots. While the scheme of pointing used can have an impact on the General Engagement measure, the other measures are unlikely to be severely affected as long as the robots point at the correct area.

Concerning the non-social presence measures, Trust, General Engagement, Cooperation, Enjoyment and Items Change, while they can be used to evaluate the use of the robotic tele-presence on a practical approach, as used by Adalgeirsson and Breazeal (2010), they can also be used to estimate the difference in cognitive presence. Since cognitive presence is involved with how the distributed group can discuss issues and cooperate in order to achieve gains in effectiveness as a group, these metrics can be used to partially evaluate the level of cognitive presence and also that of the group awareness.

One of the most affected measure between the conditions is the Item Change measure, this measure is a metric of the level of persuasion being exhibited by the remote individual being represented by the robots and their actions. According to Petty (1977), there can be two different routes through which persuasion can occur, the central and the peripheral routes. In the central route, the merit of the opposing argument is reflected on and persuasion occurs due to the argument used having an effect on the state of the individual. The peripheral route, the change in point of view is not based on the value of the opposing argument but in the manner with which it is presented to the individual. A parallel can be drawn between those routes and the cognitive and affective factors of attitudinal change (Petty & Cacioppo, 1986).

Considering the effect of these factors on the persuasion ability of an individual it was important to evaluate their contribution to the experimental task. The task was setup with guidelines for the argument used by the remote participant to respond to the choice of the local participant. This uniformity allowed for the effect from the cognitive aspect of the argument to be minimised by the design of the task, while the remaining contributing factor, the affective aspect was represented by the robots and their behaviours. Taking into account that very similar schemes were used for most of the experiment, the form and size of the robot may also be contributing factors to the affective aspect of the argument. While further work needs to be carried out in order to ascertain the exact effect of the form and size of the robot versus the behaviour of it, the use of the robots in this experiment tend to point towards the morphology argument.

Nevertheless, the integrated control scheme used in conjunction with the humanoid robot was quite successful at conveying the presence of the remote participant while still allowing the task to be experienced in a positive way by the local participants. While the enjoyment and cooperative engagement may be affected by the nature of the task the fact that there are noticeable differences for the other measures also between the different condition leads to design consideration for the robot and its behaviour for tele-presence robotic applications.

7.4.1 Summary

In this part of the investigation, it was shown that it is possible with the use of a robotic embodiment to establish a high level of presence in tele-presence settings. The robot, in this case, is used to represent the actions of a remote individual while audio communication between the members of the group is also supported. This high level of presence was obtained when a humanoid robot was used as the embodiment, and when a non-humanoid robot was used for the same purpose and using a similar control scheme, a relatively lower level of presence was obtained. When this presence was decomposed into its constituent subtypes of presence a greater insight was obtained into how the presence is impacted by the use of the robots. When taking into account the social presence exhibited by the different robots it was shown that for the humanoid robot condition, the results obtained were similar to the gold standard of social presence, the face to face, co-located, setting. While this was not the case in the non-humanoid condition, where significant differences were found between it and the face to face setting and also between the two robot conditions. From these results, two conclusions can be drawn. In the first place, with respect to the hypotheses, hypothesis H 7.2 is partially supported as the results from the humanoid robot condition have shown that it was possible to obtain a level of social presence that is comparable to a face to face setting, though this does not apply to every robot. Secondly, when analysing the individual factors of the social presence of the embodiment, it was found that hypothesis H 7.1 is also supported by the results as there was significant differences between the two robot conditions.

Moreover, when comparing the individual factors further insights were obtained. While hypothesis H 7.1 is strongly supported overall, there was one factor of social presence, Empathy, that deviated from the trend. For this factor, the impact was not limited to hypothesis H 7.1 but also had an effect on hypothesis H 7.2. While for the Empathy factor, even in the values obtained for the face to face condition, the values were generally lower when compared to the other factors, the values for both robots were significantly different from the face to face condition but were not different between the two robot conditions. While a possible reason has been put forward for the reason behind this factor being different than the others, further experiments will be needed in order to confirm this explanation.

Since the desert survival task was designed to evaluate the social presence of participants engaged in the task, social presence is an important element of the evaluation of the tele-presence through robot. However, using the dependent measures to evaluate the effectiveness of an embodiment in providing a presence, goes beyond the scope of only social presence. While social presence measures make up three of the variables, the dependent measures in comprised of eight different measures, with the other ones aimed at evaluating group collaboration. Since there was a significant difference between the two robot conditions according to social presence, it was expected that the dependent measures also show difference between the two conditions. This was confirmed by statistically different values with respect to the conditions, which supports hypotheses H 7.3, H 7.4, H 7.5, H 7.6, and H 7.7. As with the social presence, further investigation is required to confirm the factor that is behind the difference between the conditions.

7.5 Chapter Conclusion

For this chapter, the focus was on the effect of robot mediated communication on the level of presence being exhibited by a remotely located individual, whose actions are being portrayed by different robotic embodiments. As part of the investigation, an integrated scheme was developed that combined the two previous schemes, the pointing generation scheme and the gesture modification schemes, into one that would allow robots to not only portray a large set of gestures but in the case of pointing gestures, being able to point to the correct object. This scheme was implemented on two different robots, a humanoid and a non-humanoid one, which acted as the conditions for the experiment. Since the focus is on the level of presence exhibited by the embodying of the actions through robots, the task that was selected for the experiment used a version of the Desert Survival task, a task developed for assessing the level of social presence in this group task solving. The data from the experiment was captured mainly through the use of questionnaires to gauge the overall feeling that the use of each embodiment had on the participants. Due to the nature of the set task, a portion of the questionnaire was dedicated to the capture of the level of perceived social presence from the remote participant. However, since the overall presence was being assessed, other parts of the questionnaire were aimed on finding the level of cognitive presence in the two person group. After an analysis of the data it was found that there were marked differences in the level of presence being displayed in the different conditions. When the humanoid robot was used with the integrated scheme, the embodiment exhibited a level of social presence that was on par with a co-located experience while in the non-humanoid condition the level of social presence was different from both the humanoid condition and the co-located set. This difference between the conditions also extended to the other measures, where cognitive presence was also evaluated. It is to be noted that the level of empathy being exhibited by the robots were relatively low compared to the co-located setting and also while different causes have been proposed for the difference in the presence levels, further tests are required to evaluate them.

Chapter 8 Conclusions and Further Work

8.1 Chapter Introduction

In the previous chapters, an investigation into the use of robotic avatars in order to enhance tele-presence experiences was presented. In order to provide a context for this thesis, a description of the most relevant works was provided. Part of the description was focussed in identifying issues that would be investigated later in the thesis. This was followed by a description of the main hardware components of the system that was used in the thesis. Having identified both the main issues and the available hardware resources for the system, the investigation was started. This investigation being focussed on using avatars for tele-presence was involved with how the form of the avatar impacts the tele-presence experience. Since tele-presence embodiments can exhibit different types of presence, notably physical, cognitive and social (Garrison, Anderson, & Archer, 2001), the investigation was used to assess how the form of an embodiment affects the experience along each of those lines. While investigating these topics, a few side contributions were also made.

In this chapter, a brief overview of the work carried out as part of investigation is going to be presented with respect to each of the subtopics involved. Within each of these subsections, the findings of the investigation are going to be presented with respect to both the particular subtopic and also the validation of the main hypotheses presented in Chapter 2. In addition to that, this chapter is also going to focus on further avenues of research along the lines of the investigation which were not covered in this thesis but which nevertheless warrant further investigation.

8.2 Conclusions

In this section, an overview of the conclusions of the different parts of the investigation is provided. While the conclusions are mostly related to the main hypotheses, stated in Chapter 2, other findings that were also discovered during the investigation are also included.

8.2.1 Regarding Robot Pointing

In Chapter 4, the focus of the investigation was on deictic gestures, pointing gestures, being carried out by a robot. Pointing gestures are used quite frequently in human-to-human interaction, whenever there is a need to designate an object or a direction (Goodwin, 2003). While they can be as subtle as a shift in gaze, they can also be performed using hand gestures or even with other objects not innate to the body such as tools (Goodwin, 2003). Since pointing gestures form an important element of human-to-human interaction, and as a first step into testing hypothesis H 2.1, which is involved with how a local embodiment of a remote user impacts the group dynamics, the impact of autonomous pointing from robots was investigated.

Consequently two sub-hypotheses, H 4.1 and H 4.2, were formulated according to which the use of a robotic embodiment to portray the pointing may improve artefact designation; this was tested through a comparison of the times required to identify objects being designated. As part of the investigation a control module was implemented that would allow the humanoid robot to point in a human-like manner. With this new control module implemented, another sub-hypothesis, H 4.3, was also formulated according to which the human-like pointing from the humanoid would be on par with a human performing the pointing.

The experiment was conducted with respect to artefact indication under different condition. While artefact indication was limited to pointing to objects present on the tabletop interactive device, the different experiment conditions used different types of designators, including human and robots. In each condition, participants were tasked with observing the pointing actions being carried out and recording the object that was pointed to. From the results obtained from the investigation it was found that the use of a robot plays a significant role in the perception of pointing and that with an appropriate control scheme a humanoid robot can be on par with a human being in the transmission of accurate pointing information to an observer.

The results of the experiment showed that the use of a humanoid embodiment, in combination with the control scheme, significantly improved the speed with which the item being designated was identified by the participants, this was however not the case when a non-humanoid was used. From these results, hypothesis H 4.1 was partially validated and therefore some support for hypothesis H 2.1 was achieved, with respect to object identification from pointing. Hypothesis H 4.2 was also partially validated since there was an absence of significant difference between the humanoid robot condition and the co-located human condition in terms of accuracy. From the results, hypothesis H 4.3, was validated whereby a humanoid robot together with the control scheme was proved to be able to produce pointing gestures that are human-like enough, this despite the robot being restricted in its ability to replicate human pointing exactly. Having shown that humanoid robots, with the appropriate control scheme, are able to produce pointing gestures that are comparable to human deictic gestures an answer is obtained for one of the questions from section 1.3.4, question Q1.

Since this control scheme allows successful pointing generation while still being computationally cheap, it was again used in the later chapters when generation of accurate pointing gestures was required of the humanoid robot. Apart from this joint coordination scheme, another contribution from that chapter and that was used in later chapters was involved with tabletop interaction; It was found that, irrespective of the embodiment used for producing the pointing, human beings are better at finding the coordinate along one dimension than another. That is, human beings are better at identifying the column component of an object than the row component. This also means that more objects can be stacked next to each other than along the depth and still have a similar recognition rate.

8.2.2 Regarding Robot Mediated Cooperation

In Chapter 5, the focus of the investigation was on collaboration between the different members of a group when part of the group is remotely located. As pointed out by Bos et al., while collaboration between the different members of a group is facilitated when all the members of the groups are co-located, this is not the case when part of the group is remotely situated (Bos, Shami, Olson, Cheshin, & Nan, 2004). In fact, Bos et al. found that the lack of a perceived presence for the remote members of the group created a segregation of the group whereby interaction with the local members were favoured at the expense of that with the remote members (Bos, Shami, Olson, Cheshin, & Nan, 2004). Since the lack of a representation for remote participants can have a negative impact on group collaboration and with respect to hypotheses H 2.1 and H 2.2, an experiment using robots as local representations of remote participants was carried out.

As a first step into the experiment, a preliminary study was carried out with co-located groups of participants, to provide a baseline with which the results of the different experiment conditions could be compared to. Using data captured during the preliminary study a sub-hypothesis was also devised. First with respect to hypothesis H 2.1, a sub-hypothesis, H 5.1, was formulated where the use of robot embodiment was expected to provide significant improvements in task performance compared to no embodiment. An additional sub-hypothesis, H 5.3, was also formulated according to which the humanoid robot, in conjunction with the control scheme developed in Chapter 4, would allow the group to perform significantly better than the non-humanoid one. Furthermore, using the baseline for co-located obtained from the preliminary study, and with respect to hypothesis H 2.2, another sub-hypothesis, H 5.2, was formulated, that stated that the use of robot embodiments would allow the geo-separated group to collaborate similarly to it being co-located. A final sub-hypothesis, H 5.4, was formulated according to which the robot embodiments would allow a sufficiently useful representation that the same person would be leader as when they were co-located.

This experiment was carried out with a task focussed approach, where the task was a simple shuffle puzzle implemented on the tabletop interactive device that the participants could interact with by either actively moving tiles using touch or by pointing to the tiles to refer to them in conjunction with speech. This experiment was run under three conditions, no local representation, representation using a humanoid robot and representation using a non-humanoid robot. All three conditions also included the use of audio communication between the different locations. For the conditions where robot representations were used, pointing gestures were used by the robot to portray the actions of the remote participant. The pointing schemes for both robots were the same that were used in Chapter 4.

From the results obtained from that chapter, sub-hypotheses H 5.1 and H 5.3, which were concerned with performance of the group, were partially validated, this in turn supports hypothesis H 2.1. This was represented in the data as a significant difference between the humanoid robot condition and the no representation condition, in part of the task performance metrics used, while no significant difference was found between the non-humanoid robot condition and the no representation condition. Furthermore, the results also supported sub-hypotheses H 5.2 and H 5.4, although H 5.2 was only partially supported, nevertheless this also provided support to hypothesis H 2.2

Finally, in Chapter 5, answers to questions Q2 and Q3, see section 1.3.4, were obtained where it was found that a robot embodiment allows the group to have both a higher performance and also allow the group to behave in a manner more similar to when the members are co-located than when no representation were used. This higher performance is indicative of a higher cognitive presence and since the robots also establish a physical presence, a higher level of presence is obtained from a robot than when no local embodiment is used.

8.2.3 Regarding Robot Gesturing

In Chapter 6, the focus of the investigation was on gesturing, when it is being reproduced by a robot, based on gestures that have been previously captured from a human performer. As has been shown by Adalgeirsson and Breazeal (2010), gestures play an important part in establishing a presence via a robot. Moreover, Riek et al., have shown that a humanoid robot is able to successfully convey gestures that were programmed into it (Riek, et al., 2010). However, the success of having a humanoid robot, controlled using a motion capture system, at producing comprehensible gestures had yet to be determined. It should be noted that the gestures, while being able to be performed by the robot, had to be transformed, and scaled, in order to be aligned with co-occurring speech, before they could be portrayed by the robot.

Consequently, hypothesis H 6.1 was formulated according to which the recognised transformed gestures would not be significantly different from the original one. Prior to transforming the gestures, the gestures had to first be captured and therefore a relatively simple and easily moved capture system was implemented. The captured gestures were then transformed using an implemented computationally cheap gesture modification scheme. Similarly to the previous chapters, a task, requiring the participants to identify the gestures being reproduced by a humanoid, was used to evaluate the successfulness of the robot in portraying the gestures.

From the results of the presented user study, it was shown that the gestures even when being scaled to fit with the co-occurring speech were successfully identified, supporting hypothesis H 6.1. With the gesture modification scheme, used to scale the gestures, being proved to be successful, at producing identifiable gestures, it was possible to use a wider range of gestures that go beyond mere deictic gestures. Finally, one important limitation with respect to deictic gestures was also identified, whereby while the modified deictic gesture could be successfully identified the object they were designating may not be the same object that was being designated by the original gesture.

8.2.4 Regarding Robot Presence

In Chapter 7, the focus of the investigation was on the presence exhibited by robot embodiments, both humanoid and non-humanoid, tasked with the portrayal of the actions and gestures of a remote operator. It should be noted that special focus was drawn on social presence since it is the type of presence by which communication media are evaluated and have dedicated task developed to evaluating it (Lafferty, Eady, & Pond, 1970). As part of the work carried out to evaluate social presence from robot avatars, it was found that gestures play a significant part in it (Adalgeirsson & Breazeal, 2010). Since control schemes had been implemented in the previous chapters that allowed robots to successfully portray gestures, a more in-depth evaluation of hypothesis H 2.2 was performed.

Since H 2.2 is concerned with a gesture enabled tele-operated robot providing a level of presence that would allow the geo-separated group to perform on par with being co-located, a sub-hypothesis, H 7.2, was formulated focussed on social presence. Additionally, considering that while the use of gesture was evaluated but, limited focus has been drawn on how the form of a robot affects its social presence as an avatar in a tele-presence setting, other sub-hypotheses, H 7.1, H 7.3, H 7.4, H 7.5, H 7.6, and H 7.7, were also formulated.

Prior to investigating the impact of the form of the robot, acting on behalf of a remote user, on the perceived social presence it was important to be able to portray both the actions and gestures of the remote person correctly. Using the pointing generation scheme, from Chapter 4, together with the gesture modification scheme, from Chapter 6, both general gestures and pointing gestures could be performed correctly. This integrated scheme allowed for the limitation of the gesture modification scheme with respect to deictic gestures to be removed by the use of pointing scheme. Similar to the gesture capture system, from Chapter 6, a virtual tabletop interactive device system was implemented in this chapter to allow the gestures of the remote participant to be captured. This part of the investigation also involved a specific task, a Desert Survival task (Lafferty, Eady,

& Pond, 1970), that was used as part of the evaluation of the social presence exhibited by the different robot avatars.

The results obtained from that chapter provided support for all the sub-hypotheses, H 7.2, H 7.1, H 7.3, H 7.4, H 7.5, H 7.6, and H 7.7, and consequently also H 2.2. It is to be noted that the support for H 7.2 was only at a social presence level since the individual factors could only be evaluated with caution. From the results, it was also shown that the form of the robot embodiment plays a significant role on the level of presence exhibited, that is, the humanoid robot exhibited a higher social presence.

Finally, an answer was provided for question Q4, see section 1.3.4, where it was found that a humanoid robot is able to portray a higher level of presence than a non-humanoid robot and the perception of the embodied user is therefore better. Apart from providing answers to the questions from section 1.3.4, the work carried out in this thesis, have proved that both hypotheses, H 2.1 and H 2.2, are valid within the scope of this thesis. That is, the use of an embodiment especially a humanoid one using appropriate control schemes provides noticeable improvements in the exhibited presence and enhances the tele-presence experience.

8.3 Further Work

In the previous section, the work that was carried out in the investigation was described together with the important findings and the conclusions, as they relate to the main hypotheses. In this section, the other topics that were outside the scope of this thesis and those that may lead to future avenues of research are described.

8.3.1 Form Effect

One of the first avenues of possible future research is into the impact of the form of the robots on the level on interaction, with focus on the features of the robots. In this thesis, the focus was mainly on the presence and use of robots as avatars for remote members in distributed group interaction settings. During the investigation two types of robots, humanoid and non-humanoid, were used as avatars. While the results have shown that the robots, together with their behaviours, have different effects on the interaction in general; as a first step into more in-depth research into the form of the robot, studies could be carried out into which features on the robots provide a more contributing effect.

While the humanoid robot used in the experiments, had features that allowed it to be useful for the investigation, the robot itself was not as humanlike in appearance as possible. As was discussed in Chapter 3, there are a number of humanoid robots and some of these robots are even more humanlike than the RoboThespian. Moreover, some humanoid robots such as the Actroid model, apart from looking more lifelike, also possess features that allow facial expressions to be conveyed. Other humanoid robots with back-projected face screens can also transmit the actual facial expression of a remote operator. Using robots with these additional features may possibly yield even better results that the ones obtained from the humanoid robot. Consequently, one of first avenues of further research is a more fine-grained investigation into how the use of individual features from a robot, presumably humanoid, impact the presence for the same task.

8.3.2 Using Facial Expressions

Another possible future research is the use of facial expressions. During this investigation, a lot of focus was directed towards the use of gestures in combination with the form of the robot. However, the gestures being portrayed by the robots consisted mainly of hand and arm gestures. This was done since these gestures could be reproduced using either a humanoid or a non-humanoid robot and so the impact of the form of the robot could be evaluated. Using these types of gestures in combination with a humanoid robot some very positive results were obtained with respect to the level of presence being exhibited by the robot. Nevertheless, it is suggested that the level of the presence, especially with respect to social presence, can still be improved.

One of the factors of social presence is empathy, and empathy is usually best transferred through facial expressions (Cole, 2001). While there has been attempts to create empathy through robots (Kim, Kwak, & Kwak, 2009), using empathy in conjunction with a full body humanoid robot has yet to reach its full potential. As a first step in the research, a suitable representation for the face of the remote operator will have to be found. The idea, used by Adalgeirsson and Breazeal (2010), of simply using a video feed of the face of the remote operator on the head part of the robot may not be as useful on a full body humanoid robot due to the creepiness factor.

8.3.3 Duplex Robot Mediated Communication

Another avenue of research, that would expand on the work carried out in the investigation, is the use of robots for all the participants engaging in the interaction. While the findings obtained in this thesis were positive, the interaction was limited to a one-sided perception of the other member of the group. That is, the only robot representation was with respect to the remote operator. This is analogous to simplex communication. The proposed expansion would be analogous to duplex communication as each participant would be represented by a robot in each other's local environment. From the results obtained in this thesis, with respect to social presence, this expansion when implemented using humanoid robots is expected to further improve on group collaboration.

8.3.4 Object Manipulation

A further avenue of research, with respect to the work carried out in this thesis, is object manipulation. During the investigation, robots have been used to interact with objects but the objects, that the robots were interacting with, were mostly virtual ones. While these objects may have been represented on a real tabletop interactive device, all their attributes were virtual. The interactions between the robots and the objects were therefore limited to pointing to the objects with no actual touching.

One important element of interaction with real world objects is the feedback being given to the operator. This feedback can be of different types, this feedback can be limited to tactile feedback (Roke, Melhuish, Pipe, Drury, & Chorley, 2011; Roke, Spiers, Pipe, & Melhuish, 2013), if the manipulation is limited to touching, but if the manipulation involves lifting of the object a feedback, representing the change in the dynamics of the robot, will be required. While object manipulation was outside the scope of the current investigation, if it was to be performed through embodiments, it will need to support the aforementioned capabilities, in order for it to be properly represented to the operator. Since object manipulation is an important element of everyday interaction, it is a possible future avenue of research that can be explored.

8.4 Chapter Conclusion

In Chapter 1, the scope and the aim of this thesis were defined. As part of the definition of the aim, a series of objectives were identified. These objectives were in the form of questions that needed to be answered. After reviewing the literature, in Chapter 2, those questions were refined into two main hypotheses that are at the core of the thesis.

The rest of the investigation was then described starting with the main hardware components and through each of the phases of the investigation. In each phase, each described in a chapter, sub-hypotheses that relate to the main hypotheses or the questions from Chapter 1 were formulated and tested. Some of the findings of the results of the investigation were used to provide support to each of the main hypotheses from Chapter 2. From the results obtained both hypotheses were fully supported across each of the constituent aspects of presence in tele-presence.

While the humanoid robot used in this thesis, was appropriate for the work carried out, it remains a rather simple and moderately accurate robot. Therefore, the findings obtained are likely to be true and the contributions made in this thesis, with respect to the humanoid robot, able to be successfully applied to other more advanced humanoid robots.

Finally, following from the overview of the conclusions of the thesis, a section describing the potential future avenues of research was presented together with how the work carried out during this investigation fits with the future developments.

References

Adalgeirsson, S., & Breazeal, C. (2010). MeBot: A robotic platform for socially embodied telepresence. *HRI 2010* (pp. 15-22). ACM Press.

Aldebaran Robotics. (2014, November 9). *More About NAO*. Retrieved from Aldebaran Robotics Humanoid robotics & programmable robots: https://www.aldebaran.com/en/more-about

Aragon, S. R. (2003). Creating social presence in online environments. *New directions for adult and continuing education, 2003*(100), 57-68.

Augestad, K. M., & Lindsetmo, R. O. (2009). Overcoming distance: video-conferencing as a clinical and educational tool among surgeons. *World journal of surgery, 33*(7), 1356-1365.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science, 15*(6), 415-419.

Baranyi, P., & Csapo, A. (2012). Definition and Synergies of Cognitive Infocommunications. *Acta Polytechnica Hungarica, 9*(1), 67-83.

Bargh, J., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personalily and Social Psychology, 71*(2), 230.

Bartneck, C., & Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. *13th IEEE International Workshop on ROMAN .*

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes, 15*(4), 469-489.

Beciri, D. (2015, March 3). *Mobile, Dexterous, Social Robot – Nexi*. Retrieved from RobAid: http://www.robaid.com/robotics/mobile-dexterous-social-robot-nexi.htm

Benne, K. D., & Sheats, P. (1948). Functional roles of group members. *Journal of social issues, 4*(2), 41-49.

Billard, A., Robins, B., Nadel, J., & Dautenhahn, K. (2007). Building robota, a mini-humanoid robot for the rehabilitation of children with autism. *Assistive Technology, 19*(1), 37-49.

Biocca, F., Harms, C., & Gregg, J. (2001). The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. *4th annual International Workshop on Presence.* Philadelphia, PA.

Blanche, P.-A., Bablumian, A., Voorakaranam, R., Christenson, C., Lin, W., Gu, T., & Flores, D. e. (2010). Holographic three-dimensional telepresence using large-area photorefractive polymer. *Nature, 468*(7320), 80-83.

Bohren, J., Papazov, C., Burschka, D., Krieger, K., Parusel, S., Haddadin, S., . . . Whitcomb, L. (2013). A Pilot Study in Vision-Based Augmented Telemanipulation for Remote Assembly over High-Latency Networks. *ICRA.*

Booth, P. (1989). *An introduction to human-computer interaction.* Psychology Press.

Bos, N., Olson, J., Gergle, D., Olson, G., & Wright, Z. (2002). Effects of four computer-mediated communicationschannels on trust development. *CHI 2002*, (pp. 135-140).

Bos, N., Shami, N. S., Olson, J. S., Cheshin, A., & Nan, N. (2004). In-group/out-group effects in distributed teams: an experimental simulation. *CSCW 2004.* Chicago, Illinois, USA.

Boyle, E., Anderson, A., & Newlands, A. (1994). The Effects of Visibility on Dialogue and Performance in a Cooperative Problem Solving Task. *Language and Speech, 37*(1), 1-20.

Bradner, E., & Mark, G. (2002). Why distance matters: effects on cooperation, persuasion and deception. *2002 ACM conference on Computer supported cooperative work*, (pp. 226-235).

Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies, 59*(1), 119-155.

Bremner, P., Pipe, A. G., Melhuish, C., Fraser, M., & Subramanian, S. (2011). The effects of robot-performed co-verbal gesture on listener behavior. *Humanoids 2011*, (pp. 458- 465).

Broun, A., Beck, C., Pipe, T., Mirmehdi, M., & Melhuish, C. (2014). Visual Homing of an Upper Torso Humanoid Robot Using a Depth Camera. *Towards Autonomous Robotic Systems* (pp. 114-126). Springer Berlin Heidelberg.

Brown, B., & Bell, M. (2004). CSCW at play: 'there' as a collaborative virtual environment. *CSCW ‘04*, (pp. 350-359).

Butterworth, G. (2003). Pointing Is the Royal Road to Language for Babies. In S. Kita, *Pointing: Where language, culture and cognition meet* (pp. 9-34). Mahwah, Jersey: LAWRENCE ERLBAUM ASSOCIATES PUBLISHERS.

Butterworth, G., & Itakura, S. (2000). How the eyes, head and hand serve definite reference. *British Journal of Developmental Psychology, 18*(1), 25-50.

Buxton, W. (1992). Telepresence: Integrating shared task and person spaces. *Proceedings of graphics interface, 92*, 123-129.

Calinon, S., D'halluin, F., Sauser, E. L., Caldwell, D. G., & Billard, A. G. (2010). Learning and reproduction of gestures by imitation. *Robotics & Automation Magazine, IEEE, 17*(2), 44-54.

Campion, M. A., Medsker, G. J., & Higgs, A. C. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel psychology, 46*(4), 823-847.

Cassell, J. (2000). Embodied conversational interface agents. *Communications of the ACM, 43*(4), 70-78.

Cassell, J., Bickmore, T., Vilhjálmsson, H., & Yan, H. (2000). More than just a pretty face: affordances of embodiment. *5th international conference on Intelligent user interfaces*, (pp. 52-59).

Cerf, M., Harel, J., Einhauser, W., & Koch, C. (2008). Predicting human gaze using low-level saliency combined with face detection. *Advances in neural information processing systems, 20*.

Cheng, G., & Kuniyoshi, Y. (2000). Real-time mimicking of human body motion by a humanoid robot. *Sixth International Conference on Intelligent Autonomous Systems (IAS2000)*, (pp. 273-280). Venice, Italy.

Chestnutt, J., Lau, M., Cheung, G., Kuffner, J., Hodgins, J., & Kanade, T. (2005). Footstep planning for the honda asimo humanoid. *Proceedings of the 2005 IEEE International Conference on Robotics and Automation, 2005. ICRA 2005* (pp. 629-634). IEEE.

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language, 58*, 332-373.

Cohen, A. A. (1977). The Communicative Functions of Hand Illustrators. *Journal of Communication, 27*(4), 54-63.

Cohen, A. A., & Harrison, R. P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology, 28*(2), 276.

Cole, J. (2001). Empathy needs a face. *Journal of Consciousness Studies, 8*(5-7), 51-68.

Coradeschi, S., Kristoffersson, A., Loutfi, A., Von Rump, S., Cesta, A., Cortellessa, G., & Gonzalez, J. (2011). Towards a methodology for longitudinal evaluation of social robotic telepresence for elderly. *Human Robot Interaction 2011.*

Coroama, V. C., Hilty, L. M., & Birtel, M. (2012). Effects of Internet-based multiple-site conferences on greenhouse gas emissions. *Telematics and Informatics, 29*(4), 362-374.

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and cognitive neuroscience reviews, 3*(2), 71-100.

Degallier, S., Righetti, L., Natale, L., Nori, F., Metta, G., & Ijspeert, A. (2008). A modular bio-inspired architecture for movement generation for the infant-like robot iCub. *2nd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics, 2008. BioRob 2008.* (pp. 795-800). IEEE.

DiSalvo, C. F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002). All robots are not created equal: the design and perception of humanoid robot heads. *4th conference on Designing interactive systems: processes, practices, methods, and techniques*, (pp. 321-326).

Drury, J. L., Scholtz, J., & Yanco, H. A. (2003). Awareness in human-robot interactions. *Systems, Man and Cybernetics*, (pp. 912-918).

Drury, J. L., Scholtz, J., & Yanco, H. A. (2006). *Applying CSCW and HCI techniques to human-robot interaction.*

Efron, D. (1941). *Gesture and environment.* New York: King's Crown.

Egido, C. (1988). Video conferencing as a technology to support group work: a review of its failures. *In Proceedings of the 1988 ACM conference on Computer-supported cooperative work*, (pp. 13-24).

Ekman, P., & Friesen, W. V. (1981). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. In *Nonverbal communication, interaction, and gesture* (pp. 57-106).

Emery, N. J. (2000). The eye have it: The neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews, 24*, 581-604.

Engineered Arts Ltd. (2015, February 12). *Development > Hardware > Engineered Arts Ltd*. Retrieved from Engineered Arts Ltd: https://www.engineeredarts.co.uk/research/hardware/

Esenther, A., & Ryall, K. (2006). RemoteDT: Support for Multi-Site Table Collaboration. *International Conferecne on Collaboration Technologies (CollabTech).*

Fay, N., Garrod, S., & Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue: The influence of group size. *Psychological Science, 11*(6), 481-486.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*(47).

Freedman, N. (1977). Hands, words, and mind: On the structuralization of body movements during discourse and the capacity for verbal representation. In *Communicative Structures and Psychic Structures* (pp. 109-132).

Freedman, N., & Hoffman, S. P. (1967). Kinetic behavior in altered clinical states: Approach to objective analysis of motor behavior during clinical interviews. *Perceptual and motor skills, 2*(24), 527-539.

Fussell, S., Setlock, L., Parker, E., & Yang, J. (2003). Assessing the value of a cursor pointing device for remote collaboration on physical tasks. *CHI 2003.* NY: ACM Press.

Garrison, D. R. (2006). Online collaboration principles. *Journal of Asynchronous Learning Networks, 10*(1), 25-34.

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical Thinking and Computer Conferencing: A Model and Tool to Assess Cognitive Presence. *American Journal of Distance Education*.

Gaver, W. W. (1992). The affordances of media spaces for collaboration. *CSCW '92* (pp. 17-24). New York, NY, USA: ACM. doi:10.1145/143457.371596

Gera, D. L. (2003). *Ancient Greek Ideas on Speech, Language, and Civilization.* Retrieved from http://humanrobotinteraction.org/2-early-history-of-robotics-and-human-machine-interaction/

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor model. *Personality psychology in Europe, 7*, 7-28.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*(6), 516-522.

Goodrich, M., & Schultz, A. (2007). Human-Robot Interaction: A Survey. *Foundations and Trends in Human-Computer Interaction, 1*(3), 203-275.

Goodwin, C. (2003). Pointing as Situated Practice. In S. Kita (Ed.), *Pointing: Where language, culture and cognition meet* (pp. 217-241). Mahwah, Jersey: LAWRENCE ERLBAUM ASSOCIATES PUBLISHERS.

Gouaillier, D., Hugel, V., Blazevic, P., Kilner, C., Monceaux, J., Lafourcade, P., . . . Maisonnier, B. (2008). The nao humanoid: a combination of performance and affordability. *CoRR abs/0807.3223*.

Guerin, B. (1986). Mere presence effects in humans: A review. *Journal of Experimental Social Psychology, 22*(1), 38-77.

Guizzo, E. (2010). *Hiroshi Ishiguro: The Man Who Made a Copy of Himself*. Retrieved November 29, 2013, from http://spectrum.ieee.org/robotics/humanoids/hiroshi-ishiguro-the-man-who-made-a-copy-of-himself

Guo, C., Young, J. E., & Sharlin, E. (2009). Touch and toys: new techniques for interaction with a remote group of robots. *27th international conference on Human factors in computing systems*, (pp. 491-500).

Gutwin, C., & Greenberg, S. (1999). The effects of workspace awareness support on the usability of real-time distributed groupware. *Transactions on Computer-Human Interaction, 6*(3), 243-281.

Gutwin, C., Greenberg, S., & Roseman, M. (1996). Workspace awareness in real-time distributed groupware: Framework, widgets, and evaluation. In *People and Computers* (pp. 281-298).

Haddadin, S., Albu-Schäffer, A., & Hirzinger, G. (2009). Requirements for safe robots: Measurements, analysis and new insights. *The International Journal of Robotics Research, 28*(11-12), 1507-1527.

Han, J., Campbell, N., Jokinen, K., & Wilcock, G. (2012). Investigating the use of non-verbal cues in human-robot interaction with a Nao robot. *2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 679-683). Kosice, Slovakia: IEEE.

Hare, A. P. (1994). Types of Roles in Small Groups A Bit of History and a Current Perspective. *Small Group Research, 25*(3), 433-448.

Harsh, K. (2015, March 3). *NEXI – Robot with facial expressions*. Retrieved from KOTHARI HARSH: https://kothariharsh.wordpress.com/2012/05/27/nexi-robot/

Hasanuzzaman, M., Ampornaramveth, V., Zhang, T., Bhuiyan, M. A., Shirai, Y., & Ueno, H. (2004). Real-time vision-based gesture recognition for human robot interaction. *IEEE International Conference on Robotics and Biomimetics, 2004. ROBIO 2004.* (pp. 413-418). IEEE.

Heath, C. C., & Luff, P. (1991). Disembodied Conduct: Communication through Video in a Multi-Media Office Environment. *CHI '91*, (pp. 99-103).

Hewes, G. W. (1981). Pointing and language. *Advances in Psychology, 7*, 263-269.

Hindmarsh, J., Fraser, M., Heath, C., Benford, S., & Greenhalgh, C. (1998). Fragmented interaction: establishing mutual orientation in virtual environments. *CSCW '98*, (pp. 217-226). New York, NY, USA.

Hindmarsh, J., Fraser, M., Heath, C., Benford, S., & Greenhalgh, C. (2000). Object-focused interaction in collaborative virtual environments. *ACM Transactions on Computer-Human Interaction, 3*(4), 477-509.

Hirai, K., Hirose, M., Haikawa, Y., & Takenaka, T. (1998). The development of Honda humanoid robot. *1998 IEEE International Conference on Robotics and Automation*, *2*, pp. 1321-1326.

Honda. (2015, February 14). *ASIMO Specifications*. Retrieved from ASIMO Innovations by Honda: http://asimo.honda.com/asimo-specs/

Honda. (2015, February 14). *History of ASIMO Robotics*. Retrieved from ASIMO Innovations by Honda: http://asimo.honda.com/asimo-history/

Hossen Mamode, H. Z., Bremner, P., Pipe, T., & Carse, B. (2012). Cooperative Tabletop working for Humans and Humanoid Robots: Early Investigations into Artifact Indication. *3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012)*, (pp. 231-236). Kosice, Slovakia.

Hu, H., Yu, L., Tsui, P. W., & Zhou, Q. (2001). Internet-based robotic systems for teleoperation. *Int. J. Assem. Automat., 21*(2), 143–151.

Hung, H., Jayagopi, D., Yeo, C., Friedland, G., Ba, S., Odobez, J. M., . . . Gatica-Perez, D. (2007). Using audio and video features to classify the most dominant person in a group meeting. *ACM Multimedia, 3*(4).

Ickes, W. J. (1997). *Empathic accuracy.* Guilford Press.

Imai, M., Kanda, T., Ono, T., Ishiguro, H., & Mase, K. (2002). Robot Mediated Round Table: Analysis of the Effect of Robot's Gaze. *ROMAN '02*, (pp. 411-416).

International Telecommunication Union. (2003). *Transmission Systems And Media, Digital Systems and Networks.*

Isbister, K., Nakanishi, H., Ishida, T., & Nass, C. (2000). Helper agent: designing an assistant for human-human interaction in a virtual meeting space. *SIGCHI conference on Human factors in computing systems*, (pp. 57-64).

ISO. (2006). *ISO 10218-1:Robots and robotic devices - Safety requirements for industrial robots - Part 1: Robots.*

Ito, A., Hayakawa, S., Hotokata, K., & Terada, K. (2003). Can we feel a gaze pressure from a robot? Development of an eye-contact robot. *IEEE International Workshop on Robot and Human Interactive Communication*, (pp. 104-108). Millbrae, California, USA.

Jokinen, K., & Wilcock, G. (2014). Multimodal open-domain conversations with the Nao robot. *Natural Interaction with Robots, Knowbots and Smartphones*, 213-224.

Judge, T. K., & Neustaedter, C. (2010). Sharing conversation and sharing life: video conferencing in the home. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 655-658).

Kauff, P., & Schreer, O. (2002). An immersive 3D video-conferencing system using shared virtual team user environments. *Proceedings of the 4th international conference on Collaborative virtual environments* (pp. 105-112). ACM.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. *The relationship of verbal and nonverbal communication, 25*, 207-227.

Kendon, A. (2004). *Gesture: Visible Action As Utterance.* Cambridge University Press.

Kendon, A., & Laura, V. (2003). Pointing by Hand in"Neapolitan". In S. Kita, *Pointing: Where language, culture, and cognition meet* (pp. 109-137). Mahwah, Jersey: LAWRENCE ERLBAUM ASSOCIATES PUBLISHERS.

Kidd, C., & Breazeal, C. (2004). Effect of a Robot on User Perceptions. *IEEE/RSJ International Conference on Intelligent Robots and Systems.*

Kim, E. H., Kwak, S. S., & Kwak, Y. K. (2009). Can robotic emotional expressions induce a human to empathize with a robot? *18th IEEE International Symposium on Robot and Human Interactive Communication*, (pp. 358-362).

Kirk, D., Rodden, T., & Fraser, D. S. (2007). Turn it this way: grounding collaborative action with remote gestures. *SIGCHI conference on Human Factors in Computing Systems.* San Jose, California, USA: ACM.

Kita, S. (2003). Interplay of Gaze, Hand, Torso Orientation, and Language in Pointing. In S. Kita, *Pointing: Where language, culture and cognition meet* (pp. 307-328). Mahwah, Jersey: LAWRENCE ERLBAUM ASSOCIATES PUBLISHERS.

Kita, S. (2003). Pointing: A Foundational Building Block of Human Communication. In S. Kita, *Pointing: Where language, culture and cognition meet* (pp. 1-8). Mahwah, Jersey: LAWRENCE ERLBAUM ASSOCIATES PUBLISHERS.

Kleinke, C. (1986). Gaze and Eye Contact: A Research Review. *Psychological Bulletin, 100*(1), 78–100.

Kokoro Dreams. (2015, February 21). *Robot Rental.* Retrieved from Kokoro Dreams: http://www.kokoro-dreams.co.jp/rt_rent/download/act/actroid.pdf

Kose-Bagci, H., Ferrari, E., Dautenhahn, K., Syrdal, D. S., & Nehaniv, C. L. (2009). Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot. *Advanced Robotics, 23*(14), 1951-1996.

Kumotek Engineers. (2015, February 18). *[R3] Life Size Service Robot Robovie R3*. Retrieved from Humanoid & Human Interactive Robot Workshop: http://robotworkshop-kumotek.blogspot.co.uk/2011/03/r3-life-size-service-robot-robovie-r3.html

Kuzuoka, H., Oyama, S., Yamazaki, K., Suzuki, K., & Mitsuishi, M. (2000). GestureMan: a mobile robot that embodies a remote instructor's actions. *CSCW 2000*, (pp. 155-162).

Lafferty, J. C., Eady, P. M., & Pond, A. (1970). *The Desert Survival Situation.* Chicago, Illinois: Human Synergistics.

Lee, K. (2004). Presence, explicated. *Communication Theory, 14* (1), 27-50.

Lemme, A., Freire, A., Barreto, G., & Steil, J. (2013). Kinesthetic teaching of visuomotor coordination for pointing by the humanoid robot icub. *Neurocomputing, 112*, 179-188.

Levelt, W. J., Richardson, G., & La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language, 24*(2), 133-164.

Lincoln, P., Welch, G., Nashel, A., Ilie, A., State, A., & Fuchs, H. (2009). Animatronic shader lamps avatars. *8th IEEE International Symposium on Mixed and Augmented Reality*, (pp. 27-33).

Liu, P., Laffey, J. M., & Cox, K. R. (2008). Operationalization of technology use and cooperation in CSCW. *CSCW '08*, (pp. 505-514).

Lombard, M., & Jones, M. T. (2006). Defining presence. In *Immersed in media I: Telepresence theory, measurement and technology.* New York: Routledge.

Lu, Z., Lallee, S., Tikhanoff, V., & Dominey, P. F. (2012). Bent leg walking gait design for humanoid robotic child-iCub based on key state switching control. *2012 IEEE Symposium on Robotics and Applications (ISRA)* (pp. 992-998). IEEE.

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies, 7*(3), 297-337.

MacDorman, K. F., & Minato, T. (2005). *CogSci-2005 Workshop: Toward Social Mechanisms of Android Science.* Retrieved November 5, 2011, from http://www.androidscience.com/theuncannyvalley/proceedings2005/uncannyvalley.html

Masataka, N. (2003). From Index-Finger Extension to Index-Finger Pointing: Ontogenesis of Pointing in Preverbal Infants. In S. Kita, *Pointing: Where language, culture and cognition meet* (pp. 69-84). Mahwah, Jersey: LAWRENCE ERLBAUM ASSOCIATES PUBLISHERS.

Mast, M. S. (2002). Dominance as expressed and inferred through speaking time. *Human Communication Research, 28*(3), 420-450.

Matsumoto, Y., & Zelinsky, A. (2000). An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement. *IEEE Int. Conf. on Face and Gesture Recognition*, (pp. 499-505).

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*(92), 350-371.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* University of Chicago Press.

Menzel, E. W. (1974). A group of young chimpanzees in aone-acre field. *Behavior of Nonhuman Primates*, 83-153.

Metta, G., Sandini, G., Vernon, D., Natale, L., & Nori, F. (2008). The iCub humanoid robot: an open platform for research in embodied cognition. *Proceedings of the 8th workshop on performance metrics for intelligent systems* (pp. 50-56). ACM.

Microsoft. (2007, May 31). *Microsoft Surface Fact Sheet.* Retrieved March 31, 2009, from http://www.microsoft.com/presspass/presskits/surfacecomputing/docs/MSSurfaceFS.doc

Morales, A., Asfour, T., Azad, P., Knoop, S., & Dillmann, R. (2006). Integrated grasp planning and visual object localization for a humanoid robot with five-fingered hands. *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (pp. 5663-5668).

Mori, M. (1970). The Uncanny Valley. *Energy, 7*(4), 33-35.

Mutlu, B., Forlizzi, J., & Hodgins, J. (2006). A storytelling robot: Modeling and evaluation of human-like gaze behavior . *6th IEEE-RAS International Conference onHumanoid Robots, 2006* (pp. 518-523). IEEE.

Mutlu, B., Osman, S., Forlizzi, J., Hodgins, J., & Kiesler, S. (2006). Perceptions of ASIMO: an exploration on co-operation and competition with humans and humanoid robots. *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction* (pp. 351-352). ACM.

Namera, K., Takasugi, S., Takano, K., Yamamoto, T., & Miyake, Y. (2008). Timing control of utterance and body motion in human-robot interaction. *The 17th IEEE International Symposium onRobot and Human Interactive Communication, 2008. RO-MAN 2008* (pp. 119-123). IEEE.

Neto, P., Pires, J. N., & Moreira, A. P. (2010). High-level programming and control for industrial robotics: using a hand-held accelerometer-based input device for gesture and posture recognition. *Industrial Robot: An International Journal, 37*(2), 137-147.

Nguyen, D. T., & Canny, J. (2007). Multiview: improving trust in group video conferencing through spatial faithfulness. *SIGCHI conference on Human factors in computing systems* (pp. 1465-1474). ACM.

Norman, D. A. (1988). *The psychology of everyday things.* Basic books.

Oberman, L. M., Pineda, ,. J., & Ramachandran, V. S. (2007). The human mirror neuron system: A link between action observation and social skills. *Social cognitive and affective neuroscience, 2*(1), 62-66.

Oberman, L. M., Winkielman, P., & Ramachandran, V. S. (2007). Face to face: Blocking facial mimicry can selectively impair recognition of emotional expressions. *Social neuroscience, 2*(3-4), 167-178.

Ortmaier, T., & Hirzinger, G. (2000). Cartesian control issues for minimally invasive robot surgery. *2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000). 1*, pp. 565-571. IEEE.

Osgood, C., & Tannenbaum, P. (1955). The principle of congruity in the prediction of attitude change. *Psychology Review*(62), 42-55.

Parise, S., Kiesler, S., Sproull, L., & Waters, K. (1999). Cooperating with life-like interface agents. *Computers in Human Behavior, 15*(2), 123-142.

Petty, R. E. (1977). *A cognitive response analysis of the temporal persistence of attitude changes induced by persuasive communications.*

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in experimental social psychology, 19*, 123-205.

Pianesi, F., Zancanaro, M., Lepri, B., & Cappelletti, A. (2007). A multimodal annotated corpus of consensus decision making meetings. *Language Resources and Evaluation, 41*(3-4), 409-429.

Rama, J., & Bishop, J. (2006). A survey and comparison of CSCW groupware applications. *2006 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*, (pp. 198-205).

Riek, L. D., Rabinowitch, T., Bremner, P., Pipe, A. G., Fraser, M., & Robinson, P. (2010). Cooperative gestures: effective signaling for humanoid robots. *5th ACM/IEEE International Conference on Human-Robot Interaction.*

Roke, C., Melhuish, C., Pipe, T., Drury, D., & Chorley, C. (2011). Deformation-based tactile feedback using a biologically-inspired sensor and a modified display. In S. B. Heidelberg (Ed.), *Towards Autonomous Robotic Systems*, (pp. 114-124).

Roke, C., Spiers, A., Pipe, T., & Melhuish, C. (2013). The effects of laterotactile information on lump localization through a teletaction system. *World Haptics Conference*, (pp. 365-370).

Rosenberg, L. B. (1995). Virtual haptic overlays enhance performance in telepresence tasks. In *Photonics for Industrial Applications* (pp. 99-108). International Society for Optics and Photonics.

Rothwell, J. (2012). *In mixed company: Communicating in small groups.* Cengage Learning.

Rothwell, J. D. (2001). *In mixed company: Small group communication.* Harcourt College Publishers.

Rothwell, W. J. (2010). *Effective succession planning: Ensuring leadership continuity and building talent from within.* AMACOM Div American Mgmt Assn.

Sakagami, Y., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N., & Fujimura, K. (2002). The intelligent ASIMO: System overview and integration. *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2002. 3*, pp. 2478-2483. IEEE.

Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H., & Hagita, N. (2007). Android as a telecommunication medium with a human-like presence. *2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, (pp. 193-200).

Sauppé, A., & Mutlu, B. (2014). Robot deictics: How gesture and context shape referential communication. *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 342-349). ACM.

Seo, Y., Park, H., & Yang, H. S. (2003). Wearable Telepresence System using Multi-modal Communication with Humanoid Robot. *ICAT*.

Shah, J., Wiken, J., Williams, B., & Breazeal, C. (2011). Improved human-robot team performance using chaski, a human-inspired plan execution system. *Proceedings of the 6th international conference on Human-robot interaction* (pp. 29-36). ACM.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). In *Biometrika* (pp. 591-611).

Shiotani, S., Tomonaka, T., Kemmotsu, K., Asano, S., Oonishi, K., & Hiura, R. (2006). World's first full-fledged communication robot "Wakamaru" capable of living with family and supporting persons. *Mitsubishi Juko Giho , 43*(1), 44-45.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications.* London: John Wiley & Sons, Ltd.

Sirkin, D., & Ju, W. (2012). Consistency in physical and on-screen action improves perceptions of telepresence robots. *seventh annual ACM/IEEE international conference on Human-Robot Interaction.*

Skype. (2014, October 8). *What are P2P communications?* Retrieved from Skype: https://support.skype.com/en/faq/FA10983/what-are-p2p-communications

Stacey, E. (2002). Social presence online: Networking learners at a distance. In *Networking the Learner* (pp. 39-48). Springer .

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of communication, 42*(4), 73-93.

Strom, J., Slavov, G., & Chown, E. (2010). Omnidirectional walking using zmp and preview control for the nao humanoid robot. *RoboCup 2009: robot soccer world cup XIII* (pp. 378-389). Springer Berlin Heidelberg.

Takayama, L., Groom, V., & Nass, C. (2009). I'm sorry, Dave: i'm afraid i won't do that: social aspects of human-agent conflict. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2099-2108). ACM.

Takemae, Y., Otsuka, K., & Mukawa, N. (2004). An analysis of speakers' gaze behavior for automatic addressee identification in multiparty conversation and its application to video editing. In Robot and Human Interactive Communication. *ROMAN*, (pp. 581-586).

Tan, K. H., Robinson, I. N., Culbertson, B., & Apostolopoulos, J. (2010). Enabling genuine eye contact and accurate gaze in remote collaboration. *ICME.*

Tang, A., Boyle, M., & Greenberg, S. (2004). Display and Presence Disparity in Mixed Presence Groupware. (pp. 73-82). Dunedin, New Zealand: Australian Computer Society, Inc.

Tang, A., Neustaedter, C., & Greenberg, S. (2006). VideoArms: Embodiments for Mixed Presence Groupware. *HCI 2006*, (pp. 85-102).

Tannenbaum, S. I., Beard, R. L., & Salas, E. (1992). Team building and its influence on team effectiveness: An examination of conceptual and empirical developments. *Advances in psychology, 82*, 117-153.

Torok, A., Asboth, K. K., Honbolygo, F., & Csepe, V. (2012). Intensity dependent interaction in audiovisual integration. *CogInfoCom 2012*, (pp. 469-473).

Trevarthen, C., & Aitken, K. J. (2001). Infant intersubjectivity: Research, theory, and clinical applications. *Journal of child psychology and psychiatry, 42*(1), 3-48.

Tsagarakis, N. G., Metta, G., Sandini, G., Vernon, D., Beira, R., Becchi, F., . . . Carrozza, C. M. (2007). iCub: the design and realization of an open humanoid platform for cognitive and neuroscience research. *Advanced Robotics, 21*(10), 1151-1175.

Tsui, K. M., Desai, M., Yanco, H. A., & Uhlik, C. (2011). Telepresence robots roam the halls of my office building. *Human Robot Interaction 2011 Workshop*, *58.*

Tuddenham, P., & Robinson, P. (2007). Distributed Tabletops: Supporting Remote and Mixed-Presence Tabletop Collaboration. *TABLETOP '07*, (pp. 19-26).

Tuddenham, P., & Robinson, P. (2009). Territorial coordination and workspace awareness in remote tabletop collaboration. *SIGCHI Conference on Human Factors in Computing Systems*, (pp. 2139-2148).

United Nations Economic Commission for Europe;The International Federation of Robotics. (2002). *World Robotics 2002.* New York and Geneva: United Nations Publications.

VStone. (2015, February 15). *Robovie-R3*. Retrieved from http://www.vstone.co.jp/products/robovie_r3/spec.html

Wang, J. G., & Sung, E. (2002). Study on eye gaze estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 32*(3), 332-350.

Wheeless, L. R., & Grotz, J. (1977). The measurement of trust and its relationship to self-disclosure. *Human Communication Research, 3*(3), 250-257.

Wilkins, D. (2003). Why Pointing With the Index Finger Is Not Universal (in Sociocultural and Semiotics Terms). In S. Kita, *Pointing: Where language, culture and cognition meet* (pp. 171-216). Mahwah, Jersey: LAWRENCE ERLBAUM ASSOCIATES PUBLISHERS.

Wu, Y., & Huang, T. S. (1999). Vision-Based Gesture Recognition: A Review. In *Gesture-Based Communication in Human-Computer Interaction* (pp. 103-115). Springer.

Xitome Design. (2015, March 3). *Mobile Dextrous Social Robots*. Retrieved from Xitome Design: http://www.xitome.com/products_robots.php

Yanco, H. A., & Drury, J. (2004). Classifying human-robot interaction: an updated taxonomy. *2004 IEEE International Conference on Systems, Man and Cybernetics. 3*, pp. 2841-2846. IEEE.

Young, J. E., Hawkins, R., Sharlin, E., & Igarashi, T. (2009). Toward acceptable domestic robots: Applying insights from social psychology. *International Journal of Social Robotics, 1*(1), 95-108.

Appendix A

A.1 Gesture Questionnaire (Preliminary Study)

Subject Info

Name ________________________________

Age ____________

Sex ____________

Signature ________________________

Code ____________

Gesture Match

Sequence Number	Gesture Name
1	
2	
3	
4	
5	

A.2 Gesture Questionnaire (Main Study)

Subject Info

Name ____________________

Age ____________________

Sex ____________________

Signature ____________________

Code ____________________

Gesture Match

Gesture Name	Sequence Number 1	2	3	4	5
Beckon					
Good					
Give					
Okay					
Phone					
Point					
Shake Hand					
Success					
Turn					
Wait					

Summary ____________________

Sequence Order

Observer Order

A.3 Presence Questionnaire

(Pre Exp) Subject Knowledge & Info

Name ______________________________

Age __________

Sex __________

Signature _________________________

Code __________

About Robotics

Don't know anything							Know a lot

About Computers

Don't know anything							Know a lot

(Pre Exp) Big 5 Personality

	Very Inaccurate				Very Accurate
I am the life of the party.					
I feel little concern for others.					
I am always prepared.					
I get stressed out easily.					
I have a rich vocabulary.					
I don't talk a lot.					
I am interested in people.					
I leave my belongings around.					
I am relaxed most of the time.					
I have difficulty understanding abstract ideas.					
I feel comfortable around people.					
I insult people.					
I pay attention to details.					
I worry about things.					
I have a vivid imagination.					
I keep in the background.					
I sympathize with others' feelings.					
I make a mess of things.					
I seldom feel blue.					
I am not interested in abstract ideas.					

	Very Inaccurate					Very Accurate
I start conversations.						
I am not interested in other people's problems.						
I get chores done right away.						
I am easily disturbed.						
I have excellent ideas.						
I have little to say.						
I have a soft heart.						
I often forget to put things back in their proper place.						
I get upset easily.						
I do not have a good imagination.						
I talk to a lot of different people at parties.						
I am not really interested in others.						
I like order.						
I change my mood a lot.						
I am quick to understand things.						
I don't like to draw attention to myself.						
I take time out for others.						
I shirk my duties.						
I have frequent mood swings.						
I use difficult words.						

	Very Inaccurate	Very Accurate
I don't mind being the centre of attention.		
I feel others' emotions.		
I follow a schedule.		
I get irritated easily.		
I spend time reflecting on things.		
I am quiet around strangers.		
I make people feel at ease.		
I am exacting in my work.		
I often feel blue.		
I am full of ideas.		

(Post Exp-A) Social Presence – Rate Your Experience

	Strongly Disagree							Strongly Agree
I hardly noticed another individual.								
The other individual didn't notice me								
I was often aware of others in the environment								
Others were often aware of me in the room								
I think the other individual often felt alone								
I often felt as if I was all alone								
My opinions were clear to the other.								
The opinions of the other were clear.								
My thoughts were clear to my partner.								
The other individual's thoughts were clear to me.								
The other understood what I meant.								
I understood what the other meant.								
I sometimes pretended to pay attention to the other individual								
The other individual sometimes pretended to pay attention to me.								
The other individual paid close attention to me								
I paid close attention to the other individual.								
My partner was easily distracted when other things were going on around us.								
I was easily distracted when other things were going on around me.								
The other individual tended to ignore me.								
I tended to ignore the other individual.								

When I was happy, the other was happy.

When the other was happy, I was happy.

The other individual was influenced by my moods.

I was influenced by my partner's moods.

The other's mood did NOT affect my mood/ emotional-state.

My mood did NOT affect the other's mood/ emotional-state.

My actions were dependent on the other's actions.

The other's actions were dependent on my actions.

My behaviour was in direct response to the other's behaviour.

The behaviour of the other was a direct response to my behaviour.

What the other did affected what I did.

What I did affected what the other did.

My partner did not help me very much.

I did not help the other very much.

My partner worked with me to complete the task.

I worked with the other individual to complete the task.

The other could not act without me.

I could not act with the other.

(Post Exp-A) Trust (ITS) – Rate Your Experience

Trustworthy								Untrustworthy
Trustful								Distrustful
Confidential								Divulging
Benevolent								Exploitive
Safe								Dangerous
Candid								Deceptive
Not Deceitful								Deceitful
Straightforward								Tricky
Respectful								Disrespectful
Considerate								Inconsiderate
Honest								Dishonest
Reliable								Unreliable
Faithful								Unfaithful
Sincere								Insincere
Careful								Careless

(Post Exp-A) Overall Feeling – Rate Your Experience

Engagement Not Engaging								Highly Engaging
Cooperation Not Cooperative								Highly Cooperative
Enjoyment Not Enjoyable								Very Enjoyable

(Post Exp-B) Social Presence – Rate Your Experience

	Strongly Disagree							Strongly Agree
I hardly noticed another individual.								
The other individual didn't notice me								
I was often aware of others in the environment								
Others were often aware of me in the room								
I think the other individual often felt alone								
I often felt as if I was all alone								
My opinions were clear to the other.								
The opinions of the other were clear.								
My thoughts were clear to my partner.								
The other individual's thoughts were clear to me.								
The other understood what I meant.								
I understood what the other meant.								
I sometimes pretended to pay attention to the other individual								
The other individual sometimes pretended to pay attention to me.								
The other individual paid close attention to me								
I paid close attention to the other individual.								
My partner was easily distracted when other things were going on around us.								
I was easily distracted when other things were going on around me.								
The other individual tended to ignore me.								
I tended to ignore the other individual.								

When I was happy, the other was happy.

When the other was happy, I was happy.

The other individual was influenced by my moods.

I was influenced by my partner's moods.

The other's mood did NOT affect my mood/ emotional-state.

My mood did NOT affect the other's mood/ emotional-state.

My actions were dependent on the other's actions.

The other's actions were dependent on my actions.

My behaviour was in direct response to the other's behaviour.

The behaviour of the other was a direct response to my behaviour.

What the other did affected what I did.

What I did affected what the other did.

My partner did not help me very much.

I did not help the other very much.

My partner worked with me to complete the task.

I worked with the other individual to complete the task.

The other could not act without me.

I could not act with the other.

(Post Exp-B) Trust (ITS) – Rate Your Experience

Trustworthy								Untrustworthy
Trustful								Distrustful
Confidential								Divulging
Benevolent								Exploitive
Safe								Dangerous
Candid								Deceptive
Not Deceitful								Deceitful
Straightforward								Tricky
Respectful								Disrespectful
Considerate								Inconsiderate
Honest								Dishonest
Reliable								Unreliable
Faithful								Unfaithful
Sincere								Insincere
Careful								Careless

(Post Exp-B) Overall Feeling – Rate Your Experience

Engagement Not Engaging								Highly Engaging
Cooperation Not Cooperative								Highly Cooperative
Enjoyment Not Enjoyable								Very Enjoyable

Appendix B

B.1 Mathematical Solve of Targeting problem

$$c^2 = a^2 + b^2 - 2ab\cos\gamma \text{ ... (1)}$$

$$\frac{a}{\sin\alpha} = \frac{b}{\sin\beta} = \frac{c}{\sin\gamma} = D \text{ ... (2)}$$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \text{ ... (3)}$$

$$\cos^2\alpha + \sin^2\alpha = 1 \text{ ... (4)}$$

Since the solution to the problem is centred on finding a value for a in Figure 4.2-G, which equates to α in Figure 4.2-H and using the cosine rule above, the following equation is obtained:

$$a^2 = b^2 + c^2 - 2bc\cos\alpha \text{ ... (5)}$$

(5) can be rewritten as

$$\cos\alpha = \frac{a^2 - (b^2 + c^2)}{-2bc} \text{ ... (6)}$$

Using (2), where

$$\frac{a}{\sin\alpha} = \frac{b}{\sin\beta} \text{ ... (7)}$$

(7) can be rewritten as

$$a = \frac{b\ \sin\alpha}{\sin\beta} \text{ ... (8)}$$

Substituting (8) in (6)

$$\cos\alpha = \frac{(\frac{b\ \sin\alpha}{\sin\beta})^2 - (b^2 + c^2)}{-2bc} \text{ ... (9)}$$

(8) can be expanded into

$$\cos\alpha = \frac{\frac{b^2\ \sin^2\alpha}{\sin^2\beta} - (b^2 + c^2)}{-2bc} \text{ ... (10)}$$

(4) can be rewritten as

$$\sin^2\alpha = 1 - \cos^2\alpha \text{ ... (11)}$$

Substituting (11) in (10)

$$\cos\alpha = \frac{\frac{b^2\ (1 - \cos^2\alpha)}{\sin^2\beta} - (b^2 + c^2)}{-2bc} \text{ ... (12)}$$

Expanding and rewriting (12)

$$\cos\alpha - \frac{b^2\cos^2\alpha}{2bc\sin^2\beta} = \frac{\frac{b^2}{\sin^2\beta} - (b^2 + c^2)}{-2bc} \text{ ... (13)}$$

(13) can be rewritten as

$$\cos\alpha - \frac{b^2}{2bc\sin^2\beta}\cos^2\alpha = \frac{b^2}{-2bc\sin^2\beta} + \frac{-b^2}{-2bc} + \frac{-c^2}{-2bc} \text{ ... (14)}$$

(14) can be rewritten as

$$\frac{b^2}{2bc\sin^2\beta}\cos^2\alpha - \cos\alpha - \frac{b}{2c\sin^2\beta} + \frac{b}{2c} + \frac{c}{2b} = 0 \text{ ... (15)}$$

Assuming:

$$\cos\alpha = x \text{ ... (16)}$$

Substituting (16) into (15)

$$\frac{b^2}{2bc\sin^2\beta}x^2 - x - \frac{b}{2c\sin^2\beta} + \frac{b}{2c} + \frac{c}{2b} = 0 \text{ ... (17)}$$

Using (17) with (3)

$$x = \frac{1 \pm \sqrt{1 - 4\frac{b^2}{2bc\sin^2\beta}\left(\frac{b}{2c} + \frac{c}{2b} - \frac{b}{2c\sin^2\beta}\right)}}{2\frac{b^2}{2bc\sin^2\beta}} \text{ ... (18)}$$

Expanding and simplifying (18)

$$x = \frac{1 \pm \sqrt{1 - \frac{b^2}{c^2\sin^2\beta} - \frac{1}{\sin^2\beta} + \frac{b^2}{c^2\sin^4\beta}}}{\frac{b}{c\sin^2\beta}} \text{ ... (19)}$$

Substituting (16) in (19)

$$\cos\alpha = \frac{1 \pm \sqrt{1 - \frac{b^2}{c^2\sin^2\beta} - \frac{1}{\sin^2\beta} + \frac{b^2}{c^2\sin^4\beta}}}{\frac{b}{c\sin^2\beta}} \quad \text{... (20)}$$

Rearranging (20)

$$\boldsymbol{\alpha = \cos^{-1}\frac{1 \pm \sqrt{1 - \frac{b^2}{c^2\sin^2\beta} - \frac{1}{\sin^2\beta} + \frac{b^2}{c^2\sin^4\beta}}}{\frac{b}{c\sin^2\beta}}} \quad \text{... (21)}$$

B.2 Example of 3D transformation of Kinect Depth Image

In section 7.2.2.1, the depth image from the Kinect was mentioned, in order to better understand what it is and how it is used in conjunction with the regions, an example is used. In Figure B.2-A, an example of the depth image is shown, this image has been modified in order to allow the depth data to be visible, that is, each depth pixel was converted into an RGB pixel. The transformed depth image provides information about the spatial distribution of the environment and some basic information about objects present in the scene.

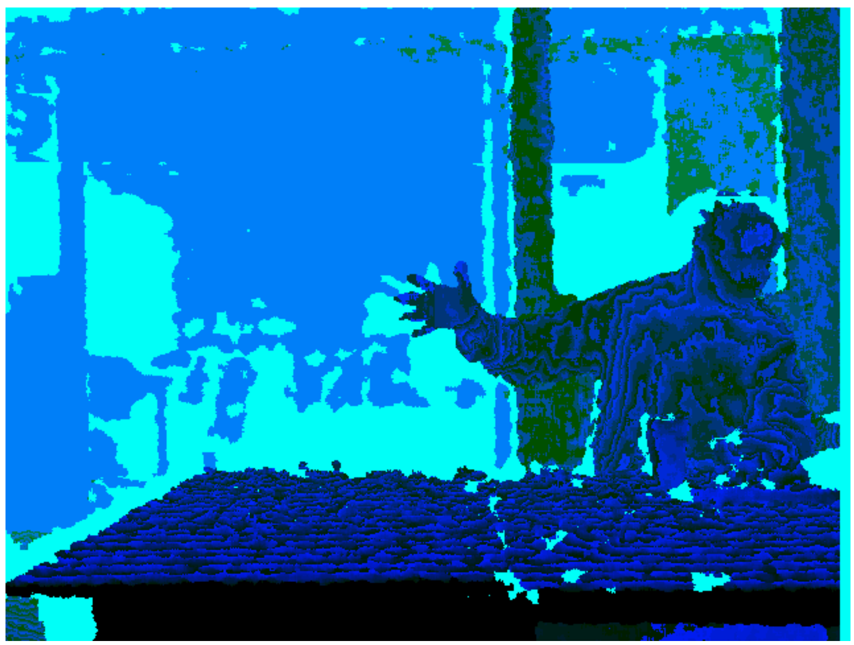

Figure B.2-A: Depth Image Example

The lighter pixels, in Figure B.2-A, represent points that are further away while the darker ones represent points that are closer. Using the data from Figure B.2-A, a reconstruction of the environment can be produced, the reconstructed environment is presented in Figure B.2-B. As can be seen the main features, from Figure B.2-A, are also present in Figure B.2-B.

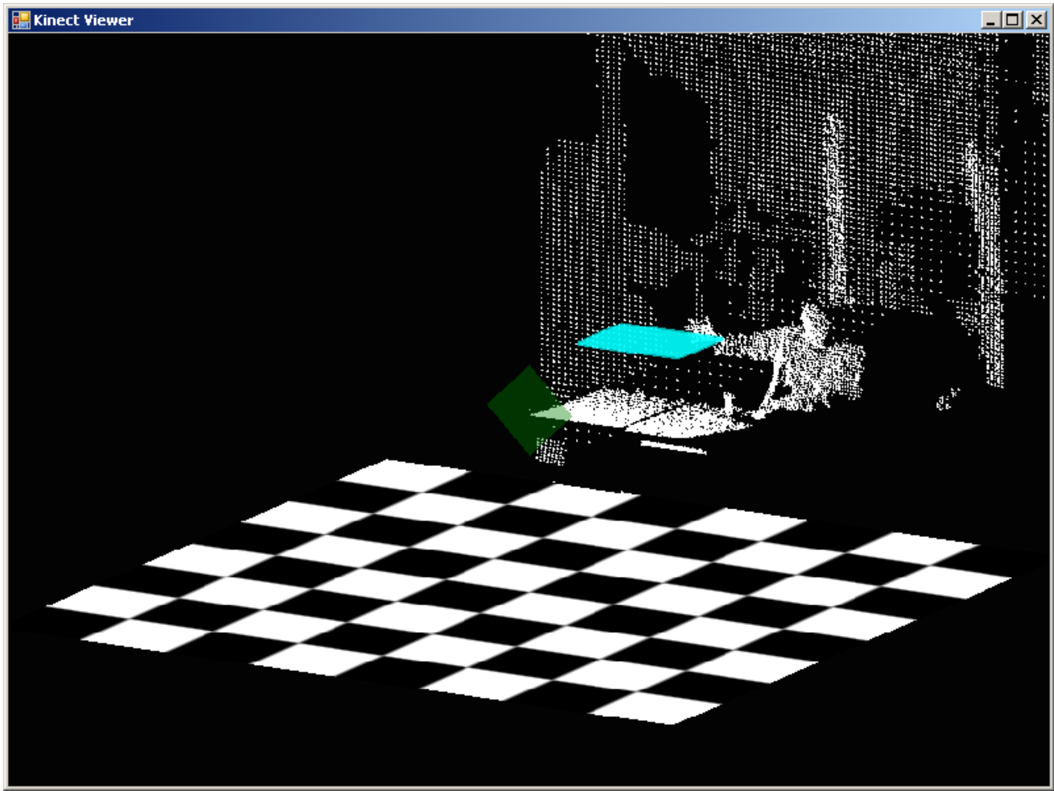

Figure B.2-B: Rendered Depth Image Example