Touch versus in-air Hand Gestures:

Evaluating the acceptance by seniors of Human-Robot Interaction using Microsoft Kinect.

by
Anouar Znagui Hassani

Thesis

for the Degree of
Master of Science

UNIVERSITY OF TWENTE.

Electrical Engineering, Mathematics and Computer Sciences

will be presented Enschede, The Netherlands

December 2011

APPROVED BY:

Dr. Betsy van Dijk	Dr. ir. Rieks op den Akker

Chair of Advisor Committee

Dr. ir. Henk Eertink	Dr. ir. Geke Ludden

Novay Enschede

Abstract

This research involves the use of an assistive robot which helps elderly people perform physical exercises. The robot presents a physical exercise on the screen which the elderly person has to copy. A camera observes the exercise performed by the elderly person. There are two ways in order to navigate though the exercises namely, in-air gestures and touch. The senior can perform a gesture or press screen buttons.

By means of an experimental comparative study, this research aims to discover among others, whether the elderly people have a preference towards an interaction modality. No significant differences were found between the interaction modalities on the technology acceptance measures on effort, ease, anxiety, performance and attitude. The scores on these measures were very high for both interaction modalities, indicating that both modalities were accepted by the elderly people. In the final interview, participants were more positive about the use of gestures than about the use of the touch modality. Most participants had a preference to use in-air gestures for the interaction with the robot because they could express themselves more using gestures as opposed to pressing touch screen buttons. An extra reason to prefer gestures were the physical constraints of many of the participants. In the touch interface they had to walk towards the robot in order to touch the screen. Of the 100 in-air gestures which are interpreted as such by the participants 93 in-air gestures were recognized as such by the gesture recognition system. Elderly participants who were unable to perform the desired gesture were disregarded in determining the quality of the gesture recognition.

Samenvatting

Dit onderzoek betreft een evaluatie van een hulprobot die senioren helpt bij het uitvoeren van lichamelijke oefeningen. De robot presenteert een oefening op het scherm die de senior vervolgens moet nabootsen. Door middel van een camera neemt de robot de gemaakte beweging waar. Er zijn twee manieren om te navigeren door de lichamelijke oefeningen, door middel van gebaren en touch. De senior kan een gebaar maken of gebruik maken van de touchscreen op de hulprobot.

Middels een experimenteel vergelijkend onderzoek is onder andere bekeken of senioren voorkeur hebben voor één van de interactievormen. Er zijn geen significante verschillen gevonden tussen de gemeten acceptatieschalen inspanning, gemak, angst, prestatie en houding. De resultaten op deze schalen waren hoog voor beide interactiemodaliteiten, wat aangeeft dat beide interactiemodaliteiten werden geaccepteerd door de senioren. In het afsluitende interview toonden de deelnemers zich positiever over het gebruik van gebaren dan over het gebruik van de touchscreen. De meeste deelnemers hadden een voorkeur voor gebaren om met de robot te communiceren, omdat de deelnemers het gevoel hadden dat zij zich beter konden uitdrukken bij het maken van gebaren in tegenstelling tot het gebruik van de touchscreen op de hulprobot. Een extra reden om gebaren te verkiezen was de fysieke beperkingen van veel van de deelnemers. Om de touchscreen te gebruiken moesten zij naar de robot lopen om het scherm te raken en dit koste in veel gevallen veel moeite. Van de 100 gebaren die de deelnemers als zodanig interpreteren, werden er 93 door het gebarenherkenningsysteem als zodanig herkend. Senioren die niet in staat waren het gewenste gebaar te maken, worden buiten beschouwing gelaten bij het bepalen van de kwaliteit van de gebaren herkenning.

Preface

This master thesis is submitted in partial fulfillment of the requirements for a Master"s Degree in computer science. It contains work done from February to September 2011. I have done my thesis at Novay in Enschede. The thesis has been made solely by me; most of the content however, is based on the research of others, and I have done my best to provide references to these sources. It all started in November 2010 when I was offered an opportunity to do my research topic course at Novay. When I finished this course I was asked to do my final thesis there as well. I didn"t hesitated a bit and accepted the offer. I would like to extend my gratitude towards people who helped me throughout. First of all I would like to thank my examination committee: Dr. ir. Rieks op den Akker, Dr. Betsy van Dijk, Dr. ir. Henk Eertink and Dr. ir. Geke Ludden.

Acknowledgments are also due to my family as they supported me and made it possible for me to pursue my education and fulfill my ambitions. Another thanks goes to my co-students for all the cooperation and friendships developed during the last years and the great times outside the studies.

Anouar Znagui Hassani

Almelo

Contents

List of Figures

List of Tables

1. INTRODUCTION

> ”Human beings try to develop machines which can make their own lives easier and richer. Robots are an example of this.” (Wadhawan, 2007)

1.1. Topic of this thesis

Humans and robots interact with each other in a variety of circumstances nowadays. Robots are performing tasks around humans within industrial and scientific settings. Their presence within the home and general society today, becomes even more common.

There is no strict definition of a "robot", but it is usually regarded as an intelligent computer which supports human goals. In recent years, another metaphor has become available: computer as an "agent". Sony AIBO, Honda"s humanoid ASIMO (Honda, 2011) and Robosoft"s Kompaï (Robosoft, 2010) are examples of advanced agents which are capable of moving, sensing their environment, performing tasks, often interacting with users via spoken natural language commands. It is also appropriate for the user to naturally interact with the robot using for example: speech, touch and/or in-air gestures. The capacity of a system to communicate with a user along different types of communication channels, and to extract and convey meaning automatically, is called multi modal interaction.

Both touch modality and in-air gestures (Fig. 2.5) are candidates for serving as modality in Human-Robot Interaction (HRI). Recent developments in technologies for the detection of in-air gestures (Kinect) have made this modality a more likely candidate than before.

This thesis presents the results of an experiment on the technology acceptance of a multimodal interactive social robot executed in a local care home called Verzorgingshuis Hoogschuilenburg (Stel, 2011). This experiment included an assistive robot which helps elderly people with performing physical exercises in a scenario called *Be Active*. The purpose of the experiment among others is to discover whether the elderly people have a preference towards an interaction modality. The work in this paper has been done at Novay for the EU FP7 project Florence[1] that focuses on personal assistive robots for Ambient Assisted

[1] http://www.florence-project.eu

Living (AAL) at home.

Fig. 1.1 shows four different kinds of robots. Sony AIBO (Fig. 1.2a) is displayed which is a *robotic pet*. Honda"s *humanoid-robot ASIMO* (Fig. 1.2d) is displayed on the right side. Kompaï (Fig. 1.2b) is an *assistive robot*, which is intended to assist the elderly in their Activities of Daily Living (ADL"s).

PeekeeII (Fig. 1.2c) has been developed by one of the partners in the Florence project called Wany Robotics as part of the Florence project. What these four robots have in common is that they all are still far from capable to naturally, adaptively and robustly interact with humans in real world situations. The current interaction modalities used in the literature involve HRI at different levels. For example recognizing in-air hand gestures (Fig. 2.6b) and facial & body posture recognition. These interaction modalities serve as a human-robot communication tool.

Figure 1.1.: Assistive Robots

(a) Sony"s AIBO **(b)** Robotsoft"s Kompaï **(c)** Wany Robotic"s PeekeeII **(d)** Honda"s ASIMO

1.2. Human-Robot Interaction Technology and User Experience

As the title of this thesis suggests, not only in-air gestures are candidates for serving as a modality in HRI. Touch modality is also a commonly used modality in e.g. mobile phones and computer screens. This study aims to discover among others, whether the elderly people have a preference towards one of the interaction modalities in-air gestures or touch. In order to measure the preference between these interaction modalities, a gesture recognition system is necessary.

Humans seem to have little difficulty in ignoring meaningless movements, while paying attention to meaningful in-air gestures. Robots or computer systems typically pay attention to all the movements, hence having great difficulty in ignoring those actions that were not intended for the system to react upon. Several terms exist for these meaningless movements (e.g. one scratching his head, rubbing his nose). Arendsen (2009) uses the term *fidgeting movements.* Fikkert (2010) identifies these non-communicative hand movements as *adaptors.*

If a gesture recognition system can ignore someone"s adaptors and positively recognize the intended gestures, then a user of that system is more likely to behave freely. Users may be able to suppress their meaningless movements, but others may be annoyed by the need to suppress part of their natural behaviour. Eventually this may lead to a restrictive experience on their physical freedom.

The knowledge gained during this study may be applied in the development of multi modal interaction systems that fit typical or natural human behaviour and capabilities.

1.3. Research questions and methods

The main question answered in this thesis is: What is the influence of multimodality in the context of HRI on user acceptance? Simply said, when an elderly person has to make use of gestures as opposed to using tactile commands to interact with the robot, does that cause differences in the user"s acceptance? This question will be further clarified in the research setup section.

The research questions were chosen because of the importance to learn more about the perception of seniors of a social robot which is equipped with multi modal interaction capabilities. The questions were:

1. Does the HRI in context of the be active scenario afford either touch or in-air gesture or both?
2. Which of the two modalities is preferred by the senior participants, or what are the objections for a particular modality against the other?
3. How would the senior participants perform a $'Next'$ and $'Previous'$ gesture without prior training?

An experiment has been performed addressing these questions involving an assistive robot. The robot presents a physical exercise on the screen which the elderly person has to copy. A camera observes the exercise performed by the elderly person. There are two ways to navigate through the exercises namely, in-air gestures and touch. The senior can perform a gesture or press screen buttons.

1.4. Outline of this thesis.

The next chapter will describe the field of multimodal interaction including the field of HRI in the context of AAL. Different studies which have been done in these fields are discussed as well.

A prototype application that was developed is discussed in Chapter 3.

Chapter 4 will outline the research question as well as how the experiment has been set up and how it is executed. Chapter 5 will present the results of the experiment after which a discussion will follow in chapter 6. Finally the conclusion and recommendations are presented in chapter 7.

The next section will briefly describe the company at which this research is conducted.

1.5. Company profile

Novay is a company that represents the development of new ways to effectuate innovation, modernization and progress, and works towards a future in which both people's personal and work lives are increasingly supported by clever ICT applications.

Novay is a participant of the EU FP7 project Florence.

Figure 1.2.: Novay

The aim of the Florence project is to improve the well-being of elderly (and that of their beloved ones) as well as improve the efficiency in care through Ambient Assisted Living (AAL) services, supported by a general-purpose mobile robot platform (Fig. 1.3). The Florence project investigates the use of such robots in delivering new kinds of AAL services to elderly persons and their care providers. The robot is the connecting element between several stand alone AAL services in a living environment as well as between the AAL services and the elderly person. Through these care, coaching and connectedness services, supported by Florence, the elderly will remain independent (Florence, 2011).

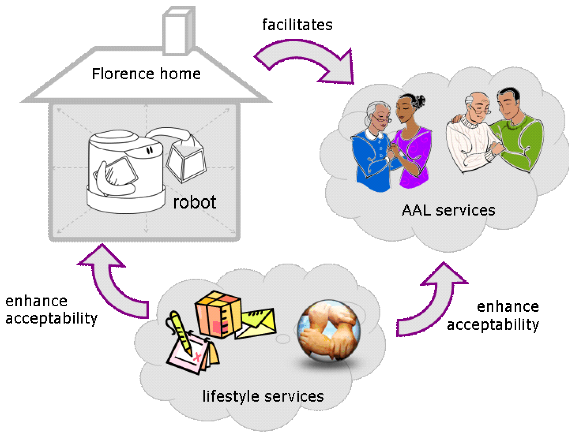


Figure 1.3.: Florence overview (Florence, 2011)

2. Related work

In the following sections, the most relevant related work in the fields of AAL and HRI are described as well as interaction modalities and the state-of-the-art technologies.

2.1. Ambient Assisted Living & Human Robot Interaction

This section will explain the concept of Ambient Assisted Living and examples of assistive technologies will be presented. The examples of assistive technologies involve robots and other computer systems which are designed to help elderly people with their Activities of Daily Living (ADL"s). This master thesis involves an assistive robot which helps elderly people perform physical exercises. The design of this robot is further described in chapter 3.

The most common medical assistive technologies such as glasses, walkers, canes and hearing devices are used in The Netherlands among adults with the age of 65 and older (Wingen, 2008) but in this thesis the main focus is set to assistive technologies targeting the specific ADL"s such as, health management and maintenance.

According to the Oxford Institute of Population Ageing (Oxford, 2011) the age composition of nearly every country is expected to move to one in which the elderly people outnumber the young. Half of the population will be aged over 50 in approximately 20 years time. Many old people need support due to the loss of mobility mainly caused by illness. Physical as well as mental activities are getting more difficult. This influences the life of the elderly people.

The discussion around Ambient Assisted Living started when political institutions could not ignore the demographic change any more. A program called AAL was started by the European Union to support the innovation of devices which maintain and improve the health of elderly people (Steg et al., 2006). Current developments include relatively simple technological devices such as an alarm button for elderly people. When the button is pressed due to a fall, it will raise an alarm to the ambulance. A more complex system includes an assistive robot which monitors and supports the activities of the daily lives of elderly persons such as the multi purpose mobile robot for AAL called Florence (Bargh and Lowet, 2010). O'Grady et al. (2010) have proposed a system with which critical situations can be detected. This system could be implemented in an assistive robot. A critical situation could be an elderly person falling in his or her home due to immobility. In their laboratory (O'Grady et al., 2010) they have multiple areas and rooms representing a fully instrumented house. Several (infrared) sensors are deployed in that house together with a multi-agent system[1]. The conditions and actions that an agent takes are encoded within the agent"s code design. Using the beliefs and rules defined within a predicate logic, agents decide how to act. See Fig. 2.1 for an illustration of the various components within such a multi-agent system for detecting critical situations.

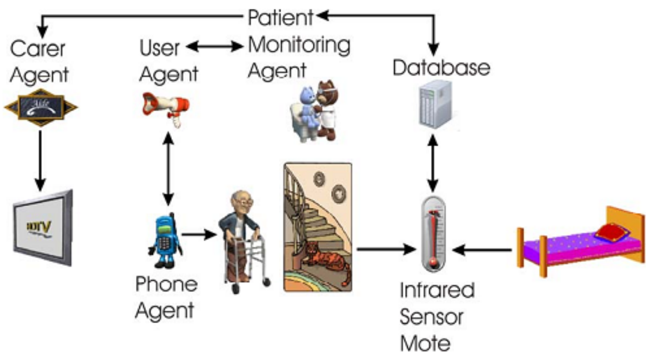


Figure 2.1.: Collaborating agents for monitoring the patient in the house (O'Grady et al., 2010)

When for example the alarm is raised because the elderly person fell down a staircase the Patient Monitoring Agent must decide what action to take. First it contacts the User Agent. As the User Agent is responsible for communicating with the patient the User Agent will first determine whether there is a visual screen in the patients vicinity to which a message can be transferred. In this case there is no visual display in the area. The User Agent will subsequently contact the patient"s Phone Agent. The Phone Agent determines that it is in the same room as the elderly person, hence a message is displayed on the mobile phone and the phone starts making an alarming sound. When the elderly person does not react within one minute. The User Agent informs the Patient Monitoring Agent that the patient has not responded to the message. The Patient Monitoring contacts the Carer Agent which is responsible for communication with the carer of the elderly person. The Carer Agent checks whether the carer is in the house through querying the database. The Carer Agent either transmits a alarm message to the visual display unit close to the

[1] A Multi-Agent Systems or MAS refers to software agents deployed in a network of computer systems. These agents are able to communicate with each other.

carer. Had the carer not been in the house, the carer"s Phone Agent would have been contacted.

An other example of an assistive technology is a robot which is called "Kompaï" (Fig. 2.2) has been developed by a company named Robosoft (Robosoft, 2010). Accordingly Kompaï is intended to help the elderly in their ADL"s. It is a mobile and communicative product, equipped with speech, it is able to understand simple orders and give a certain level of response. It knows its position within the house, how to get from one point to another on demand or on its own initiative, and it remains permanently connected to the Internet and all its associated services. Future generations of Kompaï will be equipped with visual abilities, and also the possibility to understand and express emotions.

Figure 2.2.: An assistive robot called Kompaï by Robosoft (Robosoft, 2010)

Van Breemen et al. (2005) has developed a research platform called "iCat" for studying social human-robot interaction. The platform consists of the robot character "iCat" (Fig. 2.3) iCat"s task is to recognize users, build profiles of them and handle user requests. The profiles are then used to personalize domestic functions performed by the robot e.g. different light and music conditions are used for every individual user asking iCat to create a relaxing ambiance.

Heerink et al. (2006) have summarized their experiences in collecting user data on human-robot interaction in nursing homes for the elderly. For their experiments they used the iCat and created a specific context in order for it to be used in a Wizard of Oz fashion. Elderly people were exposed to the iCat in groups of 8 participants per group.

Figure 2.3.: Human Robot Interaction with an assistive robot by Philips called: iCat (van Breemen et al., 2005)

After a short introduction the robot explained what the possibilities were: agenda-keeping, information providing or for instance companionship. A conversation with the robot took place. During the conversation the participant had to accomplish simple tasks such as setting an alarm and asking the weather forecast. The behaviour was closely monitored and recorded by camera. Learnings from two experiments were used to develop guidelines to support human-robot user studies with elderly users. The results showed that this demanded strict organization, full cooperation by nursing personnel and extreme attention to informing the participants both before and during the experiment. Moreover, analysis of the data from the studies suggests that social abilities in a robotic interface contribute to feeling comfortable talking to the robot and invite elderly people to be more expressive.

2.2. Multimodal Interactive Systems

Gibbon et al. (2000) define multimodal systems as follows:

- Multimodal systems are systems which represent and manipulate information from different human communication channels at multiple levels of abstraction.

One of the first multimodal interactive systems was Bolt"s *Put that there* - system (Bolt, 1980). With this system users could create, place and move objects in a map which was projected on the wall using gestures and speech (Fig. 2.4). Bolt"s main goal was to study how actions can disambiguate actions in another modality.

Current research which has been done in the field of multi modal interactive systems includes research by Böhme et al. (2003) who has created a multi

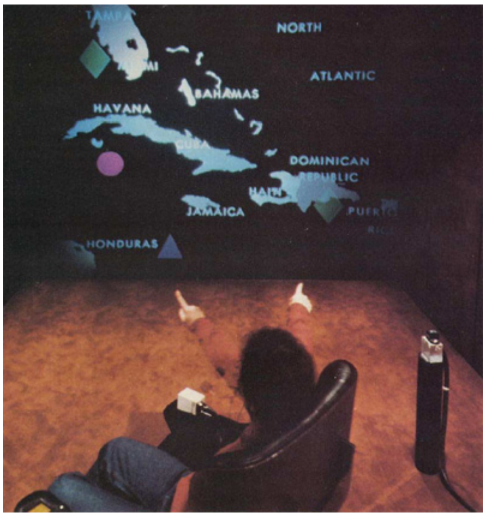


Figure 2.4.: Bolt"s *Put that there -system*

modal interaction scheme for HRI suited for service robots. During a scenario, the usage of the robot as a mobile information kiosk, methods for vision-based interaction were developed. Fong et al. (2003) did research on the notion of socially interactive robots, they discussed different forms of "social robots" which resulted in a taxonomy of design methods and system components to build an interactive social robot. Jokinen and Raike (2003) discussed multimodal technologies and how multimodal interfaces can be used to improve HRI.

Interactive robots are equipped with sensory input devices through which the robot perceives its environment. For example Kompaï (Fig. 1.2b) is fitted with a camera and several ultrasonic distance sensors to perceive its environment. A touchscreen is also present for the user as an output device, but it is also usable as an input device. According to the user manual, the robot is capable of navigating through ones home according to a given path. The robot is also capable of recognizing speech as well as speaking itself making use of a Text To Speech(TTS) system. The user is for instance able to ask *What time is it?* Due to the speech recognition system and the present dialogue manager the robot is able to respond. This specific robot makes use of the two modalities speech and touch.

2.3. Relevant Gesture Types

Another upcoming HRI modality are gestures. Although this modality might be understood by the reader, however a clear distinction has to be drawn between, what is mentioned in the title, in-air gestures and other kinds of gestures. In-air gestures like the one displayed in Fig. 2.5 are characterized by the trajectory movements of the hand. More examples are: "waving", and the gesture one would make when the term "swimming" has to be depicted.

Figure 2.5.: In-air gesture, *Come here*

Efron (1941) conducted one of the first studies of human gestures, resulting in five categories on which later taxonomies were built. The categories were physio-graphics, kinetographics, ideographics, deictics, and batons. The first two are lumped together as iconics in McNeill's classification McNeill (1992).

McNeill (1992) has identified a number of different types of gestures which people use when they interact, for example:

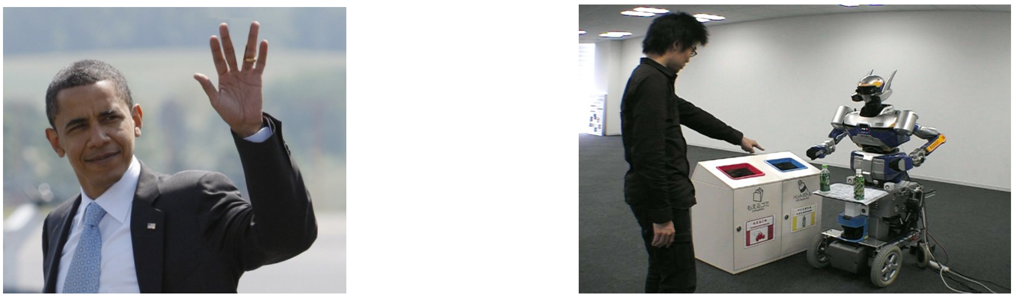
"Iconic" gestures are closely related to speech, illustrating what is being said. For example, when describing how water was poured from a glass into a dish, a child arced her fist in the air as though pouring from one container to another. See Fig. 2.6a for another example of an iconic gesture.

"Deictic" gestures have the function to suggest objects or events in a concrete world (Fig. 2.6b). These gestures are "pointing movements whose function is to indicate a concrete person, object, location, direction but also to point to unseen, abstract or imaginary things" (Krauss et al., 2000).

Only these gesture types the have been discussed in this chapter because of its relevance to this research. The gestures used in this study for example the gesture "Go to the next one" or simply "Next" (Fig. 3.5) belong to the category of deictics. Deictics are better recognized by the gesture recognition system specially built for this research. Chapter 3 will explain in more detail why deictics are better recognized than other gestures types.

2.4. Gesture Classification Procedure

A wave gesture is more difficult to recognize for a computer system than for us, human beings. Pavlovic et al. (1997) differentiates two different approaches in gesture recognition: a 3D model based and an appearance-based approach. The foremost method makes use of 3D information of key elements of the body parts in order to obtain several important parameters, like palm position or joint angles. On the other hand, appearance-based systems use images or videos for direct interpretation using for instance image processing. A trend



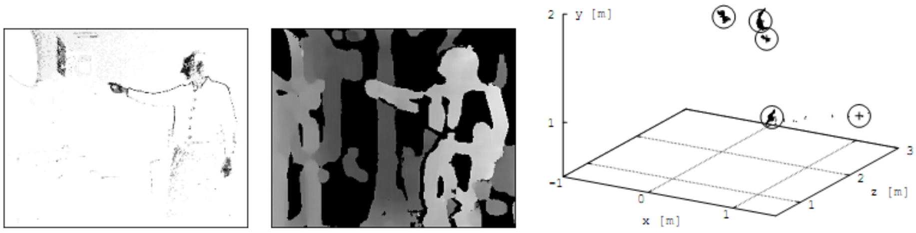
(a) Iconic Gesture, live long and prosper

(b) Deictic gesture, *I* present

Figure 2.6.: Example of gestures

is visible in current research to use a skeleton based model of the human or human parts (Pavlovic et al., 1997). Jin et al. (2011) use data-gloves in order to capture the hand to create a skeleton of the hand. The skeleton is then used to recreate a virtual model of the hand. A gesture is recognized as soon as a positive match is found comparing it with a gesture library.

Stiefelhagen et al. (2004) have built a natural multimodal HRI system which is capable of recognizing pointing gestures as well as the recognition of a person"s head orientation (Fig. 2.7).

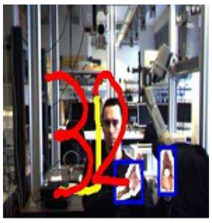


a. Skin-color map b. Disparity map c. 3D skin-color clusters

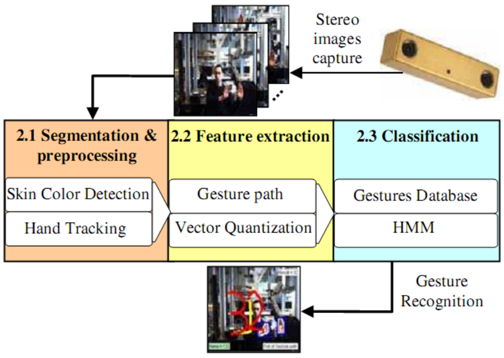
Figure 2.7.: Features for locating head and hands: Skin colored 3D pixels are clustered using K-means Algorithm. The resulting clusters are depicted by circles

Using a 3D camera, head and hands can be identified by human skin color. In combination with morphological operations it is possible to isolate the region of interest and produce closed regions. Tracking the hand consists of estimating the likelihood and compare the results against a gesture database to find a positive match. Gesture recognition is a very popular research area.

Elmezain et al. (2008) have proposed an automatic system that recognizes continuous gestures for Arabic numbers (0-9) in realtime based on Hidden Markov Model (HMM). The continuous gestures are recognized by their idea of codewords (Fig. 3.7). Their principle for computation of direction vectors is also used in the design of the gesture recognition system presented in this thesis (See Chapter 3).



(a) A movement trail of a user creating the gesture for number 32



(b) Gesture recognition process

Figure 2.8.: Gesture recognition system by Elmezain et al. (2008)

The principle works as follows: the user is located in front of a camera (Fig. 2.8). Preprocessing is done to track the hand. As the hand moves, each movement has a particular direction. The angle is computed between the previous and current location(point) as the hand moves (See Fig. 3.7a). A number (0-12) is assigned to each possible direction (See Fig. 3.7b). An example of how the number 4 would be classified is sequence of codewords 4,0,10,4,4 (Fig. 2.9). Instead of recognizing Arabic numbers, this specific method could also be used for recognizing gestures showed in Fig. 2.5 and Fig. 2.6.

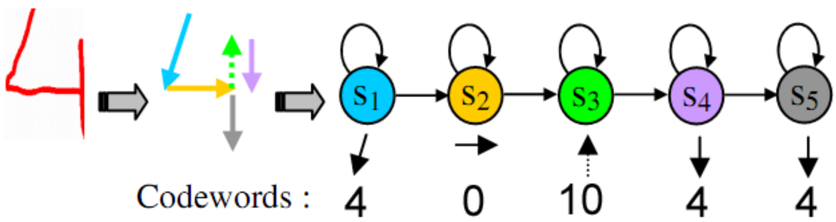


Figure 2.9.: Sequence of codewords for the number 4. This figure originates from the paper of Elmezain et al. (2008). This figure has been altered because of a different configuration of the codewords used in the design of the gesture recognition system for this research

2.5. Evaluation of Robot Acceptance in the domain of Human Robot Interaction

Relatively few studies have been performed on the acceptance of robots by elderly people in the context of assistive technology. Although the evaluation of robot acceptance seems to be one of the most important factors of getting the elderly to genuinely integrate assistive technologies in their ADL"s, it also happens to be a rather difficult subject to do research on. Several models are available to evaluate the acceptance of technological artifacts.

The first introduction of the Technology Acceptance Model (TAM) was by Davis (1989). It has become one of the most widely used theoretical models in behavioural psychology. Basically it states that Perceived Usefulness and Perceived Ease of Use determine the behavioural Intention to *Use* (Fig. 2.10) a system and the assumption exists that this behavioural intention is predicting the actual use (Taylor and Todd, 1995; Heerink et al., 2009a). The TAM is not originally developed for evaluation of Human Robot Interaction.

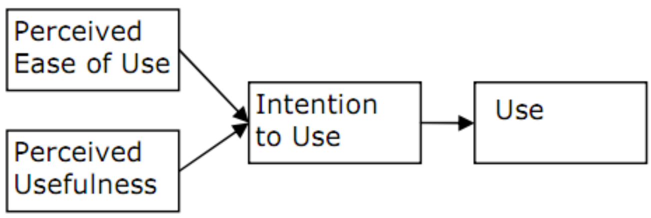


Figure 2.10.: Basic TAM assumptions (Davis, 1989).

In 2003 Venkatesh et al. (2003) have published a summation of current models and factors and presented a model called UTAUT (Unified Theory of Acceptance and Use of Technology) in which all relevant measurable factors were incorporated such as performance, effort, attitude, self-efficacy and anxiety. "Originally the TAM, related models and UTAUT were merely developed for and validated in a context of utilitarian systems in a work environment"(Heerink et al., 2009a). Heerink et al. (2009b) were the first to apply it in the Human Robot Interaction domain. Heerink et al. (2009b) have conducted experiments using the UTAUT model and they discovered that the UTAUT model had a low explanatory power in the Human Robot Interaction domain.

Also the UTAUT model introduced by Venkatesh et al. (2003) insufficiently indicated that social abilities of the robot contribute to the acceptance of a social robot (Heerink et al., 2009b). (Heerink et al., 2009b) took it a step further and extended the UTAUT model with several other constructs such as: Anxiety (ANX), Trust (Trust) Perceived Sociability (PS). See Fig. 2.11 for an overview of the complete interrelated constructs. Table 2.1 describes the definitions of each of the constructs (Heerink et al., 2009b).

Table 2.1.: Model overview

Code	Construct	Definition
ANX	Anxiety	Evoking anxious or emotional reactions when using the system.
ATT	Attitude	Positive or negative feelings about the appliance of the technology.
FC	Facilitating conditions	Objective factors in the environment that facilitate using the system.
ITU	Intention to use	The outspoken intention to use the system over a longer period in time.
PAD	Perceived adaptability	The perceived ability of the system to be adaptive to the changing needs of the user.
PENJ	Perceived enjoyment	Feelings of joy or pleasure associated by the user with the use of the system.
PEOU	Perceived ease of use	The degree to which the user believes that using the system would be free of effort
PS	Perceived sociability	The perceived ability of the system to perform sociable behavior.
PU	Perceived usefulness	The degree to which a person believes that using the system would enhance his or her daily activities
SI	Social influence	The user's perception of how people who are important to him think about him using the system
SP	Social presence	The experience of sensing a social entity when interacting with the system.
Trust	Trust	The belief that the system performs with personal integrity and reliability.
Use	Use/Usage	The actual use of the system over a longer period in time

As this study attempts to discover whether there is an influence of interaction modalities on the robot acceptance, with the UTAUT evaluation model it is possible to predict the future use of the robot acceptance on human-robot interaction. Claudine and Tinker (2005) considers the "felt need" for assistance combined with "product quality" to be the factor to evaluate the acceptance of assistive technology (Fig. 2.13). The "felt need" can be compared with the Intention To Use (ITU) construct of Heerink et al. (2009b). "Product quality" is also a factor that is considered to measure acceptance (Claudine and Tinker, 2005). "Product quality" can be related to the Perceived Ease Of Use (PEOU) construct of Heerink et al. (2009b).

In other research Heerink et al. (2009a) have conducted experiments involving the robotic agent iCat and a screen agent called Annie. They used a questionnaire in order to measure the influence of social abilities on acceptance of an interface robot and a screen agent by elderly users. The questions concerning acceptance were adapted from the UTAUT questionnaire. They adapted the questionnaire for several reasons. First some elders who piloted the questionnaire had difficulty indicating the level to which they agreed with statements and responded better to questions than to statements. Also because some participants had trouble reading, it was much easier for most of the participants if

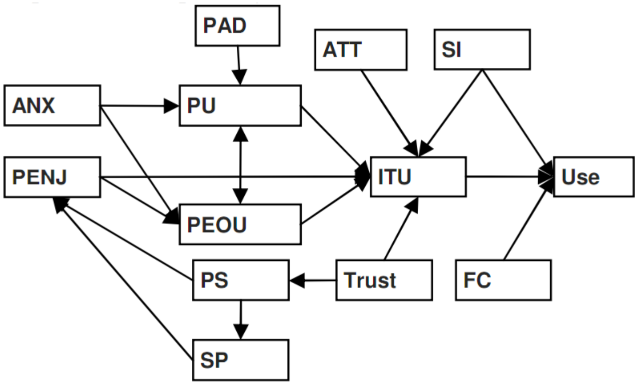


Figure 2.11.: An overview of the construct interrelations (Heerink et al., 2009b)

they were asked the questions by an interviewer who could clarify the question if necessary. Furthermore they stated that since UTAUT was developed for using technology at work, the questions needed to be adapted to a domestic user environment. The questions that could not be adapted were omitted. Finally they added five questions concerning trust and perceived social abilities.

The answers to the UTAUT questions were given on a five point scale (1 is "absolutely not", 2 is "not", etcetera). The complete questionnaire contained 27 questions of which 19 were related to UTAUT constructs. Experiments were held with a total of 42 elderly persons involving the robotic agent iCat and the screen agent Annie. The questionnaire with 27 questions was used. Comparing the results of the questionnaire regarding the robotic agent to those of the screen agent using t-tests, Heerink et al. (2009a) found no significant differences between the scores for the constructs. For the individual questions of the questionnaire they also did not find any significant differences except for one question namely if they would be afraid to make mistakes or break something $p = 0.003$. The scores for the robotic agent iCat on this particular question were much higher.

Figure 2.12.: Screen agent Annie (Heerink et al., 2009a)

Because of this difference they aimed to detect relationships among the items in the questionnaire beyond the existing constructs to be able to explore alternative constructs by detecting hidden factors which underlie the questions. After an analysis they were able to distinguish five factors. The questions of the questionnaire were regrouped according to these factors forming new constructs (Tab. 2.2). Performance and Attitude (PA) was the first construct. It measures how respondents 'see themselves' both practically and socially in the light of the new technology. They called the second construct Effort, Ease and Anxiety (EEA) which measures how easily people think they can adapt, learning how to work with the technology and overcoming eventual anxieties. Applying Cronbach's Alpha to these newly formed constructs showed that these constructs yielded an $\alpha = 0.86$ for the construct PA and $\alpha = 0.87$ for the construct EEA. Cronbach"s α (alpha) is a coefficient of reliability. It is commonly used as a measure of the internal consistency or reliability of a psychometric test score for a sample of examinees. An alpha of 0.75 indicates that the test will be 75% reliable in practice, so that the higher the Cronbach alpha, the more reliable the test results will be. A questionnaire was also designed to measure acceptance in this project using the EEA factor and the PA factor. Not the complete scale was used. Chapter 4 will discuss the used factors in more detail.

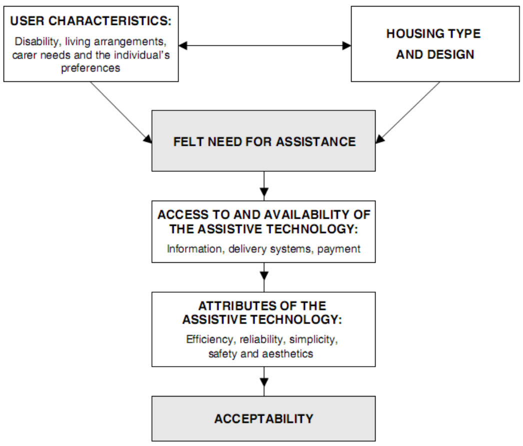


Figure 2.13.: A model of the acceptability of assistive technology by Claudine and Tinker (2005)

2.6. Conclusion

HRI, as a field, has made great strides toward understanding and improving interactions with computer-based technologies. From the early explorations of

Table 2.2.: New formed constructs by Heerink et al. (2009a)

Construct	Questions
Performance and Attitude (PA)	Do you think the system would be useful to you? Do you think the system would help you do things? Do you think the staff would be pleased if you would have this system? Do you think it is a good idea to use the system? Would you like to use the system?
Effort, Ease and Anxiety (EEA)	Do you think you can quickly learn how to control the system? Do you think the system is easy to use? Do you think you could work with the system without any help? Do you think you could work with the system if you could call someone for help? Do you think you could work with the system if you had a good manual? If you were to use the system, would you be afraid to make mistakes or break something?
Trust (TR)	Would you trust the system if it gave you advice? Would you follow the system's advice?
Sociability (SO)	As you have noticed, you control the system by speech. Do you think you can easily communicate with it that way? Would you consider the system to be social? Do you feel understood by the system?
Enjoyment (ENJ)	Did you feel uncomfortable talking to a system? Do you think many people would be pleased if you would have this system? Did you find the system a pleasant conversational partner? Do you feel at ease with the system?

direct interaction with desktop computers, we have reached the point where usability, usefulness, and an appreciation of technology's social impact, including its risks, are widely accepted. Now, advances in computer technology, artificial intelligence and speech simulation have led to breakthroughs in robotic technology that offer significant implications for the HRI. Developing a robot for elderly people which is capable of natural interaction enables cooperation and thus HRI is induced between the robot and the elderly person. Little research has been done evaluating HRI with elderly people. Especially on the evaluation of interaction modalities gestures and touch.

Several methods for measuring either social interaction or factors that have an influence on HRI have been discussed. Several subjects relevant to the research of HRI within the context of AAL have been discussed. This related section showed an interdisciplinary field of research studies ranging from assistive technologies and multimodal HRI to the social psychological approach for evaluation.

This thesis evaluates the acceptance by seniors of HRI using a service on a robot of which the design and implementation is discussed in the next chapter.

3. Design & Implementation

An application has been developed that will be used to support this study to discover whether there is a preference in interaction modality. This section will describe how the design and implementation phase is established by first explaining the scenario which will be used. A scenario has been developed in order to provide the user with a purpose to interact with the robot and to help the elderly participants stay healthy for a longer period of their life. The focus of this research is set on the evaluation of the interaction and more specific the interaction modalities touch and in-air gestures. The next subsection will provide insight on the scenario which was developed and how it is used to create the interaction.

3.1. Be Active-Scenario

A scenario has been developed whereby the elderly person performs exercises in order to improve the lifestyle of the elderly person and to stay healthy. The senior in this scenario stands in front of the assistive robot. On the screen of the robot several body postures are presented that have to be copied by the senior. After each successfully performed exercise (as detected by the detection part of the software) the senior navigates to the next or previous exercise. This is exactly the point where interaction between the elderly person and the robot is induced. The elderly user has to navigate to the next or previous exercise. This scenario enables the eldery person to interact with the robot. HRI may be realized using different modalities such as speech, head pose, gesturing and touch or a combination of these modalities. However this study compares two modalities namely touch and in-air gestures. These two modalities including the scenario were incorporated in an application. The next section will explain the design process.

3.2. Design process

The touch modality for navigation during the before mentioned scenario is relatively easy to implement as it only requires two screen buttons. One button for

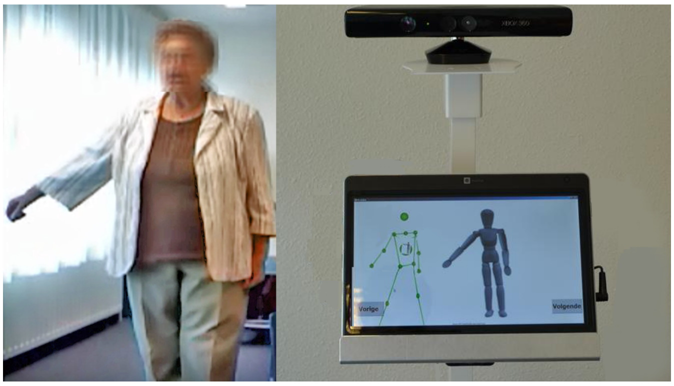


Figure 3.1.: Example of posture which is copied by a participant

navigating to the next exercise and one button for navigating to the previous exercise. The main concern regarding the design and implementation of the software application was developing a gesture recognition system in order to recognize the gestures ”*Next*” and ”*Previous*”. The reason for choosing these two modalities is the high availability of software prototyping platforms with which it is possible to design and create an application including the touch modality and the in-air gesture modality. Together with the relatively small period of time wherein this study has to be executed, these modalities are a good candidate to implement. Another reason for choosing these modalities is the fact that little research is done on the acceptance of interaction modalities involving elderly users.

In order to implement the complete system, the following components are necessary. A camera in order to observe the user. A touch screen in order to display the exercises and to receive touch commands. An exercise detection system in order to evaluate whether the physical exercises performed by the user are carried out correctly. A gesture recognition system with the purpose of recognizing the in-air gestures ”*Next*” and ”*Previous*”.

3.3. Hardware

Low budget 3D sensors are available nowadays which makes it attractive and easily accessible. Natural interaction middleware[1] handles the image processing and provides 3D points of every joint of the user"s body. Depended of which framework is chosen different Software Development Kit"s (SDK"s) exist with which an application could be developed.

Tab. 3.1 show"s the difference in traditional and contemporary research approaches regarding gesture recognition.

[1]http://www.primesense.com/

The traditional approach involves the use of a web-cam as input device. Each frame is analyzed pixel-by-pixel using various kinds of feature extraction algorithms to discover the points of interest. The contemporary approach differs from the traditional approach as preprocessing of the camera images is handled separately so that the developer is able to focus more on recognition of the gestures. The contemporary approach is used in this research.

Table 3.1.: Gesture recognition system outline

(a) Traditional

Application
Points of interest
Image Processing
Single/dual web cam

(b) Contemporary

Application
Middleware
3D sensor array

The 3D sensor which is used in this study is the Microsoft Kinect 3D sensor array (See Figure 3.2a). The Kinect sensor array exists of 2 depth-of-field sensors and an RGB camera. This 3D sensor was originally designed for the game console Xbox 360. But the 3D sensor is also usable when connected with the PC. The original PeekeeII (Robosoft, 2010) which has been mentioned in the introduction is adapted for this particular research in order to incorporate the necessary hardware parts. A stand is mounted on top of the robotic platform PeekeeII (See Figure 3.2b). A touchscreen which essentially is a touchscreen enabled laptop is mounted below the Kinect. By having the depth-of-field camera and the RGB camera a calculated distance apart, the Kinect is able to perform immediate, 3D incorporation of real objects into on-screen images.

The IR camera measures the reflected light. Due to pattern recognition on the IR points and triangulation between the source and receiver, depth is measured. PrimeSense, the company behind the technology of the Kinect talks of "LightCoding"- technology PrimeSense (2011).

3.4. Software

In order to use the 3D sensor a PC driver is installed. This is the PrimeSense Sensor driver (PrimeSense-Driver, 2011). The OpenNI (Open Natural Interaction) cross-platform framework is installed as it contains API"s for writing

(a) Microsoft Kinect

(b) PeekeeII with stand and touchscreen

Figure 3.2.: Robot platform

applications utilizing natural interaction. The application for the experiment has been written in C#.

3.4.1. Application

The application includes the interaction modalities gestures and touch. The Kinect is used to display the users body movements on the screen as well as to detect the exercises and recognize the gestures ”*Next*” or ”*Previous*”. Fig. 3.4 shows the Graphical User Interface (GUI) . Two figures are shown. The static figure on the right side shows a particular posture. More detailed: the left hand is raised and moved from the body.

The left figure shows a skeleton of the recognized body from the user standing in front of the Kinect. Two buttons, ”*Next*”and ”*Previous*” are displayed at each side of the screen to enable the touch modality. The gestures "Next" and

Figure 3.3.: IR point cloud

"Previous" are also recognized (See Fig. 3.5)
An elderly person is performing a gesture "*Next*" in Fig. 3.5. First the elderly person is standing with his both arms at each side of the body. First moving his right arm upwards. Then the elderly participant moves his hand to the right and his arm finally ends in the position it started.

3.4.2. Exercise Detection System

By using the shoulders as a reference point, the angle between the shoulder point and the hand point is calculated. The angles are continuously calculated and compared with a list containing combination of angles specifying different exercises. For example:

$$< Excersise\ String = ”ExcersiseA”\ LeftAngle = ”270”\ RightAngle\ = ”135”/ >$$

3.4.3. Gesture Recognition System

A gesture recognition system has been build using the C# framework called Accord.NET which provides many algorithms for many topics including Artificial Intelligence. It contains several methods for statistical analysis including discrete and continuous Hidden Markov Models. The left skeleton as displayed in Fig. 3.4 shows interconnected line drawing. The joints as well as the head, feet and hands are represented as dots. These dots are a representation of XYZ coordinates which are received at a rate of 30 frames per second. The following

Figure 3.4.: GUI

Figure 3.5.: Gesture "Next"

paragraphs describes the how the recognition of gestures is performed. The used techniques originate from Elmezain et al. (2008).

3.4.4. Feature extraction

Feature extraction is necessary in order to recognize the gesture path and plays a significant role in system performance. There are three basic features; location, orientation and velocity. Previous research Yoon et al. (2001) showed that the orientation feature is the best in terms of accuracy results.

Therefore it has been used as the main feature in the gesture recognition system. A gesture path is a spatio-temporal pattern which consists of points (x_{hand}, y_{hand}). The orientation is determined between two consecutive points from the hand gesture path by Eq. 3.1. were T represents the length of the gesture path. The orientation θ_t is quantized by dividing it by 20° to generate the codewords 0 to 12 (Fig. 3.7).

Each movement of the hand has a direction and because of the feature extraction algorithm, each direction has a codeword. So for instance when one makes a gesture "Slow down" like depicted in Fig. 3.6. According to Fig. 3.7 the sequence of codewords could be [9,9,9,4,4,4] or [10,10,10,4,4,4] or other combinations. The codeword sequence [9,9,9,4,4,4] can be interpreted as first going up (Fig. 3.7b) and code 4 can be related with movement downwards. All the possible combinations for this particular gesture can be stored in order to use it for evaluation explained in the next subsection.

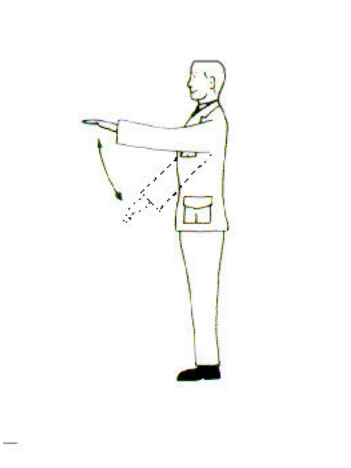


Figure 3.6.: "Slow down" gesture

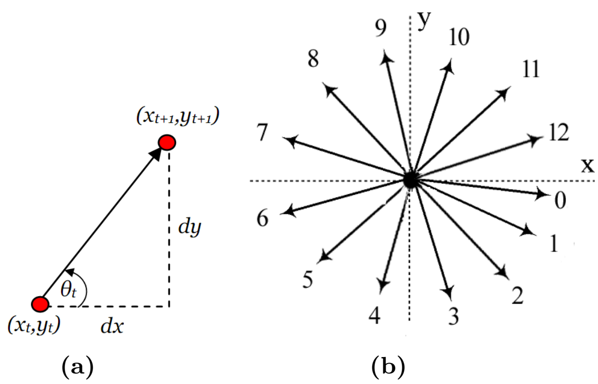


Figure 3.7.: The orientation and its codewords (a) Orientation between two consecutive points (b) directional codewords from 0 to 12 (Elmezain et al., 2008)

$$\theta_t = \left| \frac{\arctan 2\left(\frac{y_{t+1}-y_t}{xt+1-x_t}\right) \cdot 180}{\pi} \right| \qquad ; t = 1, 2, ..., T\ -\ 1 \tag{3.1}$$

3.4.5. Classification

The final stage in the gesture recognition system is classification. The gesture sequences (codeword sequences) are classified by evaluating a set of Hidden Markov Models in order to check which could have generated a given new sequence of observations (codewords). The Forward algorithm is executed in each of the models, and selects the one with highest probability. Moreover, Baum-welch algorithm is used for training to construct a gesture database. The gesture database contains 5 sequences for the gesture "$Next$" and 5 sequences for gesture "$Previous$".

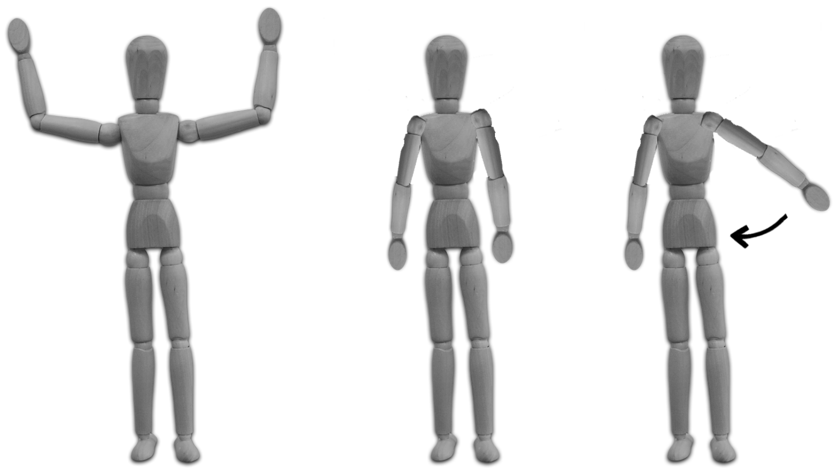
As the user performs a gesture, a sequence of codewords is observed. For instance the gesture "$Next$" may result the following sequence codewords

$[0, 1, 2, 3, 4, 5]$ and the gesture "$Previous$" may output $[5, 4, 3, 2, 1, 0]$. A pilot experiment has been performed in order to tweak several parameters of the classification system in order to improve the recognition of in-air gestures. For instance the observation length has been set to 6. And sampling time is dependent on the difference in movement. In the software this is called the update margin and has been set to 10. This means that when there is no movement or the movement is too small, the classification mechanism does not receive any observation. When the difference between the current X or Y coordinate and the previous X or Y coordinate is greater than or equal to 10 centimeter, the classification mechanism receives the observation.

3.4.6. Gesture recognition

The following concept is derived from the way sign language is performed and can be recognized by software as documented inArendsen (2009). The user usually starts from a neutral posture (Fig. 3.8a). When the user performs a gesture for the action "$Next$", the user lifts his or her right arm towards the right side of the body (Fig. 3.8b) and returns to the neutral position. The observation of the codewords generated by the return path (Fig. 3.8c) is used as input to the sequence classifier. The advantage of this approach is that, it ignores the different movements different people make as long the gesture performance occurs at the right side of the body the hand will return to the neutral position. The gesture recognition system is implemented in such a way that it even if the "Next" or "Previous" gesture is not performed precisely as suggested it can be recognized correctly. Due to the pilot study mentioned before and the tweaked parameters including the recognition of the return path, the system provides a robust gesture recognition. The gestures ”*Next*”and ”*Previous*” belong to the category of deictics as explained in Section 2.3. An important property of such a gesture is the movement from a neutral position towards the gesture itself and back towards the neutral position.

Figure 3.8.: Gestures



(a) Start pose **(b)** Neutral position **(c)** Gesture return path

3.5. Conclusion

For this comparative study between interaction modalities touch and in-air gestures a specially designed software application has been developed which is capable of recognizing trajectory movements of, in this case, a hand. State of the art technology has been applied in this study in order to provide the elderly participant with a practical and unique experience. Hence, the results of this study will be based on first hand experience of the participants and therefore valuable information may become visible. Both the application and the robot were discussed in this chapter.

For the application a scenario has been chosen in which the elderly person performs exercises in order to improve lifestyle behaviour. The senior in this scenario stands in front of the assistive robot. On the screen of the robot several body postures are presented that have to be copied by the senior. After each successfully performed posture (as recognized by the recognition part of the software) the senior navigates to the next or previous exercise. HRI may be realized using different modalities such as speech, head pose, gesturing and touch or a combination of these modalities. This study compares two modalities namely touch and gestures. The main concern regarding the design and implementation of the software application was the gesture recognition system. Having discussed the design and implementation, the next chapter will describe how the experiment has been set up.

4. Research Setup

The main question answered in this thesis is: What is the influence of multimodality in the context of HRI on user acceptance? Simply said, when an elderly person has to make use of gestures as opposed to using tactile commands to interact with the robot, does that cause differences in the user"s acceptance? This question will be further clarified in the research setup section.

The research questions were chosen because of the importance to learn more about the perception of seniors of a social robot which is equipped with multi modal interaction capabilities. The questions were:

1. Does the HRI in context of the be active scenario afford either touch or in-air gesture or both?
2. Which of the two modalities is preferred by the senior participants, or what are the objections for a particular modality against the other?
3. How would the senior participants perform a "*Next*" and "*Previous*" gesture without prior training?

An experiment has been performed addressing these questions involving an assistive robot. The robot presents a physical exercise on the screen which the elderly person has to copy. A camera observes the exercise performed by the elderly person. There are two ways to navigate through the exercises namely, in-air gestures and touch. The senior can perform a gesture or press screen buttons.

The next sections will further elaborate on the design and procedure of this research.

4.1. Design

A within subject design was chosen to measure the preference and differences between the use of the modalities gestures and touch. A within subject design has been chosen so that the participants have the possibility to choose an interaction modality based on their experience using both modalities. Counter balancing was applied to avoid order effects among the two modalities.

The extended UTAUT model of Heerink et al. (2009a) provides a questionnaire involving many factors which will determine the actual use and acceptance of a robotic system. Not all factors of that UTAUT model were used. Only the factors EEA and PA were chosen for usage in this study because the focus of this study is to evaluate the interaction between the participant and the robot and especially the usage and acceptance of the modalities gestures and touch. Also because of the Another reason for choosing only these factors was that the experiment was desired to keep as short as possible. The estimated time for one experiment was 30 minutes. The questions related to the factors EEA and PA have been altered into statements for usage in the questionnaire. These statements are believed to be more clear.

4.2. Subjects

For this study 12 elderly participants were invited to participate in this experiment. Every participant was exposed to both of the interaction modalities. These participants from a local care home called Verzorgingshuis Hoogschuilenburg in Almelo (Stel, 2011) participated voluntarily in this study, and signed a consent form for their participation (See Appendix A). The average age of the participants was 77.17 (Std. deviation: 7.19) with the youngest being 71 and the oldest 96. Of the 12 participants 7 were female. 8 participants had mobility problems. In the demographic questionnaire (See Appendix B), most of participants reported to never have used a computer before. The most frequent appliances used by the participants were the TV, coffee machine and microwave.

Figure 4.1.: Hoogschuilenburg (Carint-Reggeland, 2011)

4.3. Procedure

Each participant was welcomed in the experiment room (See Fig. 4.2). The participant started with filling in a demographic questionnaire with questions regarding for instance their daily use of appliances (See Appendix B).

After that each participant was asked whether he or she knew the definition of gestures, and how he or she would perform a "*Next*" or "*Previous*" gesture before showing how the actual gesture should be performed in order for the system to be recognized.

The participant is asked to stand in front of the robot (not asking for a specific position). The participant starts by performing a start pose (See 3.9a) in order for the system to calibrate. The user is then recognized and the first exercise is displayed for the participant to copy. When the exercise has been performed successfully, the robot will emit a voice saying "good job". Now the participant has to make clear to the robot that he or she wants to navigate to the next exercise. This is the point where either touch or in-air gestures are necessary. When the modalities are used, dependent of the action "*Next*" or "*Previous*" the robot will emit a voice saying "*Next*" or "*Previous*" as feedback towards the participant. Participants have to perform three exercises, hence each modality has to be used three times. After each modality experiment the participant was asked to sit down in order to fill in a questionnaire (See Appendix C

and D) regarding the particular modality (See Fig. 4.3).

A short interview was held at the end in order to discover what the participants found of each interaction modality, what they noticed about it and the participants were asked whether they would accept such a robot in their homes.

Figure 4.2.: Experiment room

4.4. Instruments

The preference and differences between the use of the modalities was measured using the factors Effort, Ease & Anxiety (EEA) and Performance & Attitude (PA) of the UTAUT model from Heerink et al. (2009a). Other factors such as

Social Presence (SP) Facilitating Conditions (FC) of the UTAUT model (See sec. 2.5) were omitted in this research as the focus is set on the interaction.

A questionnaire was used in combination with a 7 -point Likert scale of which the answers ranged from 1 meaning *"I absolutely disagree"* to 7 meaning *"I absolutely agree"*. Every answer can be given a number or value so that a statistical interpretation can be assessed.

Tab. 4.1 shows the questionnaire items regarding the modalities in-air gestures and touch. Not the complete scales were used. A subset of questions per scale were used to keep the experiment short. Six questions regarding the use of in-air gestures and six questions relating to the use of the touchscreen. Each question is coded and relates either to the EEA or the PA factor.

Table 4.1.: Questionnaire items Gesture and Touch

In-air gestures	Question
G_EEA_Q1	I think I can quickly learn how to communicate with the robot using gestures.
G_EEA_Q2	The gestures are easy to perform.
G_EEA_Q3	The next time I could perform the gestures without any help.
G_EEA_Q4	I get anxious when I use gestures to communicate with the robot.
G_PA_Q5	I found it pleasant to perform gestures in order to communicate with the robot.
G_PA_Q6	I have objections against performing gestures in order to communicate with the robot.
Touch	**Question**
T_EEA_Q1	I think I can quickly learn how to communicate with the robot by pressing screen buttons.
T_EEA_Q2	Pressing screen buttons is easy to perform.
T_EEA_Q3	The next time I could press the screen buttons without any help.
T_EEA_Q4	I get anxious when I press screen buttons in order to communicate with the robot.
T_PA_Q5	I found it pleasant to perform gestures in order to communicate with the robot.
T_PA_Q6	I have objections against pressing screen buttons in order to communicate with the robot.

A video recording was made of each participant during both the experiment and the complete interview (Fig. 4.3). A web cam was used. The experiment was recorded to observe afterwards whether participants understood the definition of gestures. Secondly to record the gestures which were made after the participants were asked how they would make a gesture for "Next" or "Previous".

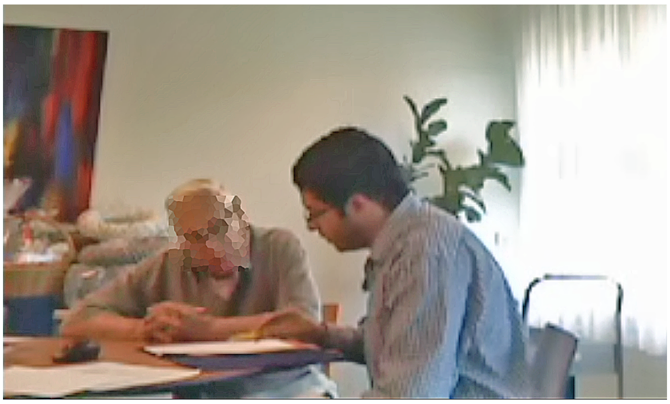


Figure 4.3.: Participant fills in the questionnaire

5. Results

This chapter presents the results of the questionnaire regarding the evaluation of interaction modalities in-air gestures and touch. The results of the interview are also present in this chapter.

One of the items in the gesture questionnaire asked whether the participants found the gestures easy to perform. A 7 -point Likert scale was used of which the answers range from "1" meaning I absolutely disagree to "7" meaning I absolutely agree. 6 out of 12 answered with the highest possible score 7 with an average score of 6.4. The exact same result was discovered after the analysis of the question regarding the touch modality wherein the question was asked whether the participant found it easy to press the screen buttons. Each participant answered to both questions about both interaction modalities (two related samples design). In this particular case the dependent variable was the factor Ease Effort and Anxiety. The independent variable was the interaction modality. An appropriate statistical test for comparing two related samples is the Wilcoxon signed-rank test. A significant difference between the related pairs is determined by $p < 0.05$. Testing these results with a Wilcoxon signed-rank test to a neutral result, yielded $p = 0.55$.

Tab. 5.1 shows the result of the Wilcoxon signed-rank test in which paired samples were used. No significant differences were found.

Table 5.1.: Statistics

Question pair	Asymp. Sig. (2 - tailed)
G_EEA_Q1 / T_EEA_Q1	0.483
G_EEA_Q2 / T_EEA_Q2	0.557
G_EEA_Q3 / T_EEA_Q3	0.569
G_EEA_Q4 / T_EEA_Q4	0.380
G_PA_Q5 / T_PA_Q5	0.589
G_PA_Q6 / T_PA_Q6	0.581

A complete questioning session with a participant took an average of 28.10 minutes (Std. deviation: 8.40). On the question *"I get anxious when I use gestures to communicate with the robot"* 8 out of 12 answered with a 7 meaning that they are not anxious performing gestures in order to communicate with the robot. In comparison with the same question for the interaction modality touch this pair yielded a $p = 0.38$.

9 out of 12 answered with a 7 on the question regarding being anxious using the interaction modality touch. Almost every participant laughed as they saw a skeleton on the screen that moved exactly the same as they did.

In Table 5.2 the average values over all the pairs for the factors EEA and PA and both interaction modalities gestures and touch are reported. Although this study reports the average values, the analysis between the different interaction modalities uses the non-parametric Wilcoxon signed-rank test, because with the low amount of pairs a normal distribution cannot be guaranteed.

Table 5.2.: An overview of all average values (Likert scale) and standard deviation within parentheses, over all the pairs for the factors EEA and PA for both interaction modalities Gestures and Touch

Factor	Gestures	Touch
EEA	6.13 (1.02)	6.18 (1.01)
PA	6.01 (1.43)	6.00 (1.84)

Section 2.5 discussed the reliability of the constructs PA and EEA ($\alpha = 0.86$ for the construct PA and $\alpha = 0.87$ for the construct EEA (Heerink et al., 2009a)). The scales PA and EEA which were used in the questionnaire of the experiment done in this research (Tab. 4.1) yielded $\alpha = 0.72$ for the construct PA and $\alpha = 0.41$ for the construct EEA. Although the used scales were not complete the construct PA with $\alpha = 0.72$ indicates that the scale is reliable ($\alpha \geq 0.7$). This is not true for the EEA construct.

The interview yielded valuable information concerning alternative gestures by the participants for the concepts"*Next*" and "*Previous*". Also interesting behavior was noticed after the preliminary question about the notion of gestures. 4 out of 12 participants knew instantly what gestures are and they even gave examples of gestures which they used back in the days during work or sports. Although they had different ideas about the performance of the gestures "$Next'$ and "*Previous*", they did not have any problems understanding and relating the specified gestures to the concepts "*Next*" and "*Previous*".

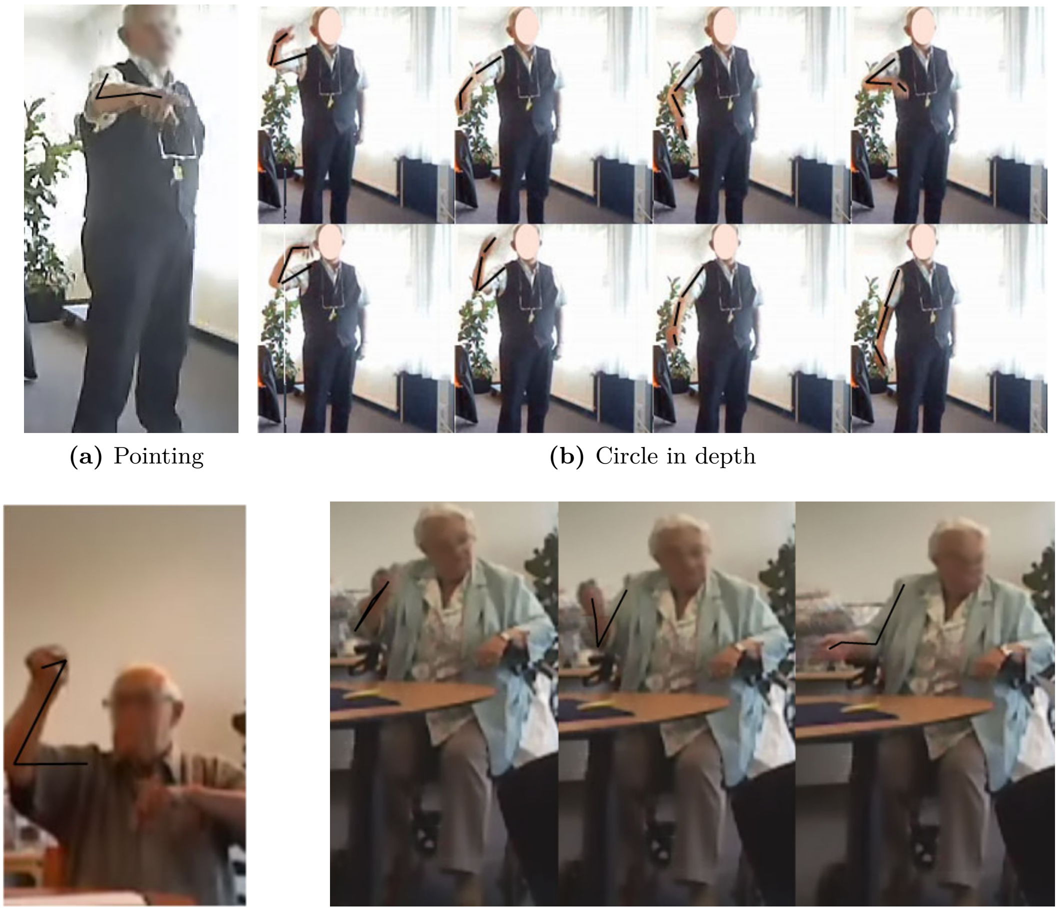
93% of the performed gestures was recognized immediately. 6% of the gestures was recognized after 2 or 3 attempts. 1% was not recognized due to physical restrictions of the participant. 7 participants used assistive technologies such as walkers and electronic wheelchairs. The optimal distance between the participant and the robot for the recognition of the in-air gestures was between 1.5 and 3 meters. The participants automatically took a position between the optimal distance.

Finally, is there a difference in performance between the gestures "*Next*" and "*Previous*"? Does the notion of "*Next*" or "*Previous*" lead to different gesture performances?

Regarding the first research question: Does the HRI in context of the be active scenario afford either touch or in-air gesture or both? 7 out of 12 participants felt that they had to make gestures to communicate with the robot. To press either one of the screen buttons, the participant had to walk towards the robot and as they weren"t mobile enough to do that they kept standing on their place and performed gestures in order to communicate with the robot. While the participants were performing the exercises, they were standing in front of the robot at a distance of more or less 1.7 meters.

The results support the second research question as well. 9 out of 12 participants argued that they could express themselves more using in-air gestures as opposed to pressing screen buttons. The final research question regarding the gesture performance resulted in different notions of the gestures "*Next*" and "*Previous*" . Fig. 5.1 shows 4 different performances of the gesture "*Next*" performed by 3 different participants. Fig. 5.2 depicts 4 different "*Previous*" gestures.

In the final interview, participants were more positive about the use of gestures than about the use of the touchscreen. 9 out of 12 participants preferred the gesture interface. Many participants (7 out of 12) argued that they could express themselves more using in-air gestures as opposed to pressing screen buttons. Physical constraints of the participants was also a cause of the before mentioned preference, as they had to walk towards the robot in order to touch the screen. They also reported that they have little knowledge about assistive robots. They were inquisitive and felt the need to have more information which is expected to result in a overall higher level of robot acceptance.



(a) Pointing **(b)** Circle in depth

(c) Pointing **(d)** Wave to the right

Figure 5.1.: Different performances of the gestures next

(a) Pointing (b) Wave to the left

(c) Both arms and hand back and forth

Figure 5.2.: Different performances of the gestures previous

6. Discussion

All definitions of a social assistive robot are built upon the same basic idea:
An effective socially assistive robot must understand and interact with its environment, exhibit social behavior, and focus its attention and communication on the user in order to help the user achieve specific goals (A.Tapus, 2006).
The assumption exists that in the near future assistive robots will be able to help the seniors to live independently. For instance robots will be able to do tasks in the household, accompany them in lonely periods or observe their health status. Studying the level of acceptance and use of assistive robots is therefore crucial, so that future assistive robots can be adjusted which is important for future adoption of robotic technology.
The acceptance and use of two interaction modalities were compared in this study; in-air gestures and touch. A scenario has been chosen in which the elderly people perform exercises in order to improve lifestyle behaviour. It was expected that due to the experiment setup, the modality preference would be in-air gestures. The experimental setup required the participant to stand in front of the robot with a distance of approximately 1.5 meters. The modality preference depends partly on the task which the user has to fulfill. For tasks that demand remote control gestures could be a better interaction modality. No significant differences were found between the interaction modalities gestures and touch.
In the following paragraphs the results are discussed according to the research questions.

Does the HRI in context of the be active scenario afford a certain type of modality touch or in-air gestures? Yes, the be active scenario in which the task was to perform exercises did afford the interaction modality in-air gestures. Analysis of the results indicated that both interaction modalities scored high on the factors Effort Ease & Anxiety and Performance & Attitude. During the interview participants were asked whether they understood what was meant by "Gestures". 8 out of 12 indicated knowing what gestures are. The remaining 4 participants understood the notion of gestures after they were told about sign language. Given examples such as waving goodbye, these 4 participants fully understood what was meant by gestures. The 12 elderly participants did not barely use modern technology other than the typical household appliances. Remarkably enough they were able to work with both interaction modalities. Apparently the interaction modalities fit well within their knowledge and maybe they will be able to handle even more complex technologies in the future.

Which of the two modalities is preferred by the senior participants, or what are the objections for a particular modality against the other? No significant differences were found between the questions regarding Effort Ease & Anxiety and Performance & Attitude. In the final interview, participants were more positive about the use of in-air gestures than about the use of the touchscreen. The modality preference depends partly on the task which the user has to fulfill. For tasks that demand remote control in-air gestures could be a better interaction modality.

Is there a difference in performance between the in-air gesture "*Next*" **and** "*Previous*"**? Does the notion of** "*Next*" **or** "*Previous*" **lead to different gesture performances?** According to Fig. 5.1 and Fig. 5.2, different gestures were performed as an answer to the question how the participants would make a gesture for "*Next*" or "*Previous*". The participants made pointing gestures for the gesture "*Next*" as well as for the gesture "*Previous*". Other types of movements were made as well, such as a circle in depth (Fig. 5.1b) and waving to the right starting with the right hand (Fig. 5.1d). The performances of the "*Previous*" gesture were different as well. A pointing gesture is showed in Fig. 5.2a as well as a wave to the left starting with the left hand. Fig. 5.2c shows a gesture made with two hands moving back and forth.

7. Recommendations & Conclusion

Since this study indicated a preference between the two modalities gestures and touch, a few recommendations for future work are in place.

Further research in this field of study is recommended to incorporate more elderly participants in the hope to gather more data in a short-term research. The design of the research is recommended to include the appearance component of the robot as well as adding speech as an interaction modality. The questioning session with the elderly participants during this research took about 30 minutes which is the recommended time frame to collect data as the intention is to not mentally overload the elderly participant with questions. A longitudinal experiment is suggested to involve a well designed robot in which the three interaction modalities touch, in-air gestures and speech are tested.

The next recommendation concerns the usage of the 3D sensor. Since the Microsoft Kinect is normally used in combination with the game console Xbox 360; computer drivers other than normal were installed on the computer laptop in order for the sensor to be useful. The drivers used in this study were unofficial, meaning that these drivers are not supported by Microsoft.

Currently Microsoft has created a special software development kit called:

The Kinect for Windows SDK beta (Microsoft-Kinect-SDK, 2011) which is a starter kit for applications developers that includes API"s, sample code, and drivers. This SDK enables the academic research to create rich experiences by using Microsoft Xbox 360 Kinect sensor technology on computers running

Windows 7. Future research is recommended to use the newly released Beta SDK of Microsoft which is better as it does not require the user to perform a certain pose in order for the system to recognize the user.

This study has not taken into account the possible influences of the robot"s appearance on elderly persons. Similar to the research of Robins et al. (2004), it is recommended to perform a study to discover the possible influences of the robot"s appearance on elderly persons. The task of the robot used in this study was to discover what the influence of multi-modality is in the context of HRI on user acceptance.

The exercises which have been performed by the elderly persons in this research included physical movements of the upper part of the body. Further research in this field is recommended to incorporate lower body exercises.

Two interaction modalities were compared; in-air gestures and touch. No significant results were found regarding the variables EEA and PA. This is possibly due to a ceiling effect. As the participants scored high on one modality, it is almost impossible to significantly score higher on the other modality. The results on these variables were very high for both interaction modalities, indicating that both modalities were accepted by the elderly people.

The results on questions in the final interview where people were asked to compare the use of the two modalities indicate that the participants reacted more positive towards the use of in-air gestures. Most participants had a preference to use in-air gestures for the interaction with the robot because they could express themselves more using gestures as opposed to pressing touch screen buttons. An extra reason to prefer gestures were the physical constraints of many of the participants. In the touch interface they had to walk towards the robot in order to touch the screen. In-air gestures can be further applied in for instance *calling* the robot, as well as *interrupting* the robots activity.

A key aspect of the Florence project is user acceptance. Florence aims to improve the acceptance of AAL (robotic) services by providing both assistance and fun oriented lifestyle services via the same means. The insights gained in this research might contribute to the integration of a multi-modal interaction system in the context of AAL, striving for full acceptance by elderly people.

The appearance might have an influence on the variable EEA of elderly people. Robins et al. (2004) have studied the influence of robots with different appearances on children. They compared the level of interaction with and response to the robot in two scenarios. The results indicated the children"s preference in their initial response for interaction with a plain, featureless robot over interaction with a human like robot.

Although no significant difference was found between the two modalities gestures and touch, participants were excited to observe how technology has improved. After the participants made the start pose almost every participant started laughing of excitement (See Figure 3.9a).

The robot used in this research is not a human like robot and the overall high level of acceptance might be the result of the appearance. It also might be possible that the featureless appearance was a salient factor causing the elderly participants to react more positive about the acceptance and interaction (this needs further investigation).

Bibliography

Arendsen, J.: 2009, 'Seeing Signs. On the appearance of manual movements in gestures'. Ph.D. thesis, Delft University of Technology.

A.Tapus: 2006, 'Towards Socially Assistive Robotics'. International Journal of the Robotics Society Of Japan.

Bargh, M. and D. Lowet: 2010, 'Florence - A Multipurpose Platform for Ambient Assisted Living'. http://www.florence-project.eu.

Böhme, H.-J., T. Wilhelm, J. Key, C. Schauer, C. Schröter, H.-M. Groß, and T. Hempel: 2003, 'An approach to multi-modal human-machine interaction for intelligent service robots'. *Robotics and Autonomous Systems* **44**(1), 83–96.

Bolt, R. A.: 1980, 'Put-that-there: Voice and gesture at the graphics interface'. *SIGGRAPH Comput. Graph.* **14**, 262–270.

Carint-Reggeland: 2011, 'Verzorgingshuis Hoog Schuilenburg'. http://www.carintreggeland.nl/.

Claudine, M. and A. Tinker: 2005, 'The acceptability of assistive technology to older people'. *Ageing and Society* **25**, 91–110.

Davis, F. D.: 1989, 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology'. *MIS Quarterly* **13**(3), pp. 319–340.

Efron, F.: 1941, 'Gesture and Environment'. *King's Crown Press, Morningside Heights, New York.*

Elmezain, M., A. Al-Hamadi, J. Appenrodt, and B. Michaelis: 2008, 'A Hidden Markov Model-Based Continuous Gesture Recognition System for Hand Motion Trajectory'. Technical report, Institute for Electronics, Signal Processing and Communications (IESK).

Fikkert, F.: 2010, 'Gesture Interaction at a Distance'. Phd thesis, Universiteit Twente.

Florence: 2011, 'Florence website'. http://www.florence-project.eu.

Fong, T., I. Nourbakhsh, and K. Dautenhahn: 2003, 'A survey of socially interactive robots'. *Robotics and Autonomous Systems* **42**(3-4), 143 – 166.

Gibbon, D., I. Mertins, and R. e. Moore: 2000, *Handbook of multimodal and spoken dialogue systems: resources, terminology.* Dordrecht: Kluwer.

Heerink, M., B. Kröse, B. Wielinga, and V. Evers: 2009a, 'Measuring the influence of social abilities on acceptance of an interface robot and a screen agent by elderly users'. In: *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology.* Swinton, UK, UK, pp. 430–439, British Computer Society.

Heerink, M., B. Kröse, V. Evers, and B. Wielinga: 2009b, 'Measuring acceptance of an assistive social robot: A suggested toolkit'. In: *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on.* p. 528.533.

Heerink, M., B. Kröse, B. Wielinga, and V. Evers: 2006, 'Human-Robot User Studies in Eldercare: Lessons Learned'. In: *Proceedings ICOST.* p. 31.38.

Honda: 2011, 'The honda humanoid robot ASIMO'. http://world.honda.com/ASIMO/index.html.

Jin, S., Y. Li, G. ming Lu, J. xun Luo, W. dong Chen, and X. xiang Zheng: 2011, 'SOM-based hand gesture recognition for virtual interactions'. In: *VR Innovation (ISVRI), 2011 IEEE International Symposium on.* pp. 317 –322.

Jokinen, K. and A. Raike: 2003, 'Multimodality - Technology, Visions and Demands for the Future'. In: *In: Proceedings of the 1 st Nordic Symposium on Multimodal Interfaces.*

Krauss, R. M., Y. Chen, and R. F. Gottesman: 2000, 'Lexical gestures and lexical access: A process model'. In: D. McNeill (ed.): *Language and Gesture*. New York: Cambridge University Press, pp. 261–283.

McNeill, D.: 1992, *Hand and Mind*. University of Chicago Press.

Microsoft-Kinect-SDK: 2011, 'Kinect for Windows SDK beta'.

O'Grady, M., C. Muldoon, M. Dragone, R. Tynan, and G. O.Hare: 2010, 'Towards evolutionary ambient assisted living systems'. *Journal of Ambient Intelligence and Humanized Computing* **1**, 15–29.

Oxford: 2011, 'Population Ageing'. Univerity of Oxford, http://www.ageing.ox.ac.uk/research.

Pavlovic, V. I., R. Sharma, and T. S. Huang: 1997, 'Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review'. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 677–695.

PrimeSense: 2011, 'Reference Design'. http://www.primesense.com/?p=514.

PrimeSense-Driver, O.: 2011, 'OpenNI Driver'. http://www.openni.org/downloadfiles/osources.

Robins, B., K. Dautenhahn, R. T. Boekhorst, and A. Billard: 2004, 'Robots as Assistive Technology - Does Appearance Matter?'. In: *Proceedings of the 13th IEEE International Workshop on Robot and Human Communication (RO-MAN 2004)*. pp. 277–282.

Robosoft: 2010, 'Robosoft introduces Kompaï'. http://www.doc-center.robosoft.com/.

Steg, H, S. H., C. Lofoff, C. Hull, and S. Schmid: 2006, 'Europe Is Facing a Demographic Challenge Ambient Assisted Living Offers Solutions'.

Stel, D.: 2011, 'Verzorgingshuis Hoog Schuilenburg'. http://www.carintreggeland.nl/.

Stiefelhagen, R., C. Fugen, R. Gieselmann, H. Holzapfel, K. Nickel, and A. Waibel: 2004, 'Natural human-robot interaction using speech, head pose and gestures'. In: *Proceedings. 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004. (IROS 2004)*. pp. 2422 – 2427 vol.3.

Taylor, S. and P. A. Todd: 1995, 'Understanding Information Technology Usage: A Test of Competing Models'. *Information Systems Research* **6**(2), 144–176.

van Breemen, A., X. Yan, and B. Meerbeek: 2005, 'iCat: an animated user-interface robot with personality'. In: *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*. New York, NY, USA, pp. 143–144, ACM.

Venkatesh, V., M. G. Morris, G. B. Davis, and F. D. Davis: 2003, 'User Acceptance of Information Technology: Toward a Unified View'. *MIS Quarterly* **27**(3), pp. 425–478.

Wadhawan, V.: 2007, 'Robots of the future'. *Resonance* **12**, 61–78. 10.1007/s12045-007-0073-7.

Wingen, M.: 2008, 'No further increase in possession medical assistive techonologies'. CBS.

Yoon, H.-S., J. Soh, Y. J. Bae, and H. S. Yang: 2001, 'Hand gesture recognition using combined features of location, angle and velocity.'. *Pattern Recognition* **34**, 1491–1501.

Appendix A Consent Form

Toestemmingsverklaring

voor deelname aan het wetenschappelijk onderzoek:

Onderzoek naar acceptatie robothulp.

1. Ik ben over het onderzoek geïnformeerd.
2. Ik ben in de gelegenheid gesteld om vragen over het onderzoek te stellen.
3. Ik heb over mijn deelname aan het onderzoek kunnen nadenken.
4. Ik stem toe met deelname aan het onderzoek.
5. Ik geef toestemming voor het bewaren van de verzamelde gegevens.
6. Ik stem toe met de video opname.
7. Ik geef toestemming de gegevens te verwerken voor het onderzoek.
8. Ik geef toestemming voor het bewaren van de verzamelde gegevens door de onderzoekers van NOVAY en Universiteit Twente.
9. Mijn persoonlijke gegevens worden niet bewaard noch gebruikt.

Naam:
Geboortedatum:…………………………………………………….
Geslacht: M / V

Handtekening: ……………………..Datum:………………….

Ondergetekende, verantwoordelijke onderzoeker, verklaart dat de hierboven genoemde persoon zowel schriftelijk als mondeling over het bovenvermelde onderzoek is geïnformeerd.

Naam: Anouar Znagui Hassani
Functie: Afstudeerder

Handtekening:……………………Datum:………………………..

Appendix B Demographic Questionnaire

Experiment vragenlijst

Bedankt voor uw deelname aan mijn afstudeerexperiment over interactie tussen mens en robot. Dit experiment bestaat uit drie delen. In het eerste deel mag u een vragenlijst invullen, zodat wij uw resultaten kunnen analyseren en vergelijken met andere participanten. Uw antwoorden blijven anoniem.

Tijdens het tweede gedeelte zult u een unieke ervaring opdoen, u mag namelijk gaan communiceren met de hulprobot door middel van het uitvoeren van gebaren en het drukken op schermknoppen. Op deze manieren kunt u de robot uitleggen wat u wilt.

Tenslotte is er een korte interview met u, waarin een aantal vragen worden gesteld over het tweede deel van het experiment.
Dit experiment zal circa 20 tot 30 minuten duren.

Deel 1

Bent u links- of rechtshandig?: Links / Rechts

Wat zijn uw hobby's?:..

Maakt u gebruik van apparaten? Zo ja welke apparaten zijn dat?

(Denkt u bijvoorbeeld aan tv, mobiele telefoon.)

....

....

....

Welke apparaten gebruikt u regelmatig?

....

....

Appendix C Questionnaire Gestures

Deel 2: Vragenlijst gebaren

1. Ik denk dat ik snel kan leren hoe ik de robot kan aansturen met gebaren.

Absoluut mee oneens	1	2	3	4	5	6	7	Absoluut mee eens

2.De gebaren zijn makkelijk uit te voeren.

Absoluut mee oneens	1	2	3	4	5	6	7	Absoluut mee eens

3. De volgende keer zou ik de gebaren kunnen uitoefenen zonder hulp.

Absoluut mee oneens	1	2	3	4	5	6	7	Absoluut mee eens

4. Ik word zenuwachtig wanneer ik gebaren maak om met de robot te communiceren.

Absoluut mee oneens	1	2	3	4	5	6	7	Absoluut mee eens

5. Ik vond het leuk om gebaren te maken en op die manier met de robot te communiceren.

Absoluut mee oneens	1	2	3	4	5	6	7	Absoluut mee eens

6. Ik heb bezwaar tegen het uitoefenen van gebaren om met de robot te communiceren.

Absoluut mee oneens	1	2	3	4	5	6	7	Absoluut mee eens

Indien van toepassing, uw bezwaar:

………..

………..

Appendix D Questionnaire Touch

Deel 2: Vragenlijst aanraakscherm

1. Ik denk dat ik snel kan leren hoe ik de robot kan aansturen door te drukken op de schermknoppen.

Absoluut mee oneens	1	2	3	4	5	6	7	Absoluut mee eens

2. Het is makkelijk om op de schermknoppen te drukken.

Absoluut mee oneens	1	2	3	4	5	6	7	Absoluut mee eens

3. De volgende keer zou ik zonder hulp de schermknoppen kunnen bedienen.

Absoluut mee oneens	1	2	3	4	5	6	7	Absoluut mee eens

4. Ik word zenuwachtig wanneer ik op de schermknoppen druk om met de robot te communiceren.

Absoluut mee oneens	1	2	3	4	5	6	7	Absoluut mee eens

5. Ik vond het leuk om op de schermknoppen te drukken en op die manier met de robot te communiceren.

Absoluut mee oneens	1	2	3	4	5	6	7	Absoluut mee eens

6. Ik heb bezwaar tegen het, drukken op schermknoppen om met de robot te communiceren.

Absoluut mee oneens	1	2	3	4	5	6	7	Absoluut mee eens

Indien van toepassing, uw bezwaar:

………..

………..

Appendix E Detailed Statistics

Question/Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
Age	79	81	71	96	78	79	84	86	89	94	88	80
G_EEA_Q1	6	7	7	7	3	7	4	6	7	7	6	6
G_EEA_Q2	7	7	7	7	6	7	6	6	7	5	6	6
G_EEA_Q3	5	6	4	6	4	6	4	7	6	6	7	6
G_EEA_Q4	6	6	7	4	5	7	7	7	7	7	7	7
G_PA_Q5	7	7	7	7	6	2	4	6	6	6	6	7
G_PA_Q6	1	1	2	1	1	2	4	1	5	1	1	1
T_EEA_Q1	7	6	6	7	7	6	4	6	6	7	6	3
T_EEA_Q2	7	5	7	7	7	6	4	7	6	6	7	6
T_EEA_Q3	6	3	7	6	7	2	5	7	7	7	7	6
T_EEA_Q4	7	7	6	7	7	6	6	7	7	7	7	7
T_PA_Q5	7	7	7	7	7	2	6	4	6	7	7	6
T_PA_Q6	1	1	1	1	1	2	2	6	7	1	1	1

Appendix F AmI 11 Paper

Touch versus In-Air Hand Gestures: Evaluating the Acceptance by Seniors of Human-Robot Interaction

Anouar Znagui Hassani[1], Betsy van Dijk[1], Geke Ludden[2], and Henk Eertink[2]

[1] Human Media Interaction, Twente University, Enschede, The Netherlands
a.znaguihassani@student.utwente.nl, e.m.a.g.vandijk@utwente.nl
[2] Novay, Enschede, The Netherlands
{henk.eertink,geke.ludden}@noyay.nl

Abstract. Do elderly people have a preference between performing in-air gestures or pressing screen buttons to interact with an assistive robot? This study attempts to provide answers to this question by measuring the level of acceptance, performance as well as knowledge of both interaction modalities during a scenario where elderly participants interacted with an assistive robot. Two interaction modalities were compared; in-air gestures and touch. A scenario has been chosen in which the elderly people perform exercises in order to improve lifestyle behavior. The seniors in this scenario stand in front of the assistive robot. The robot displays several exercises on the robot screen. After each successfully performed exercise the senior navigates to the next or previous exercise. No significant differences were found between the interaction modalities on the technology acceptance measures on effort, ease, anxiety, performance and attitude. The results on these measures were very high for both interaction modalities, indicating that both modalities were accepted by the elderly people. In a final interview participants reacted more positive on the use of in-air gestures.

Keywords: Robot Acceptance, Assistive technologies, Activities of daily Living (ADL's), Human Robot Interaction.

1 Introduction

Both touch modality and in-air gestures are candidates for serving as modality in Human-Robot Interaction (HRI). Recent developments in in-air gestures (Kinect) have made this modality a more likely candidate than before.

This thesis presents the results of an experiment on the technology acceptance of a multimodal interactive social robot. The work in this paper has been done at Novay for the EU FP7 project Florence (http://www.florence-project.eu/) that focuses on personal assistive robots for Ambient Assisted Living (AAL) at home. The research involves an experiment using an assistive robot called *Florence*[1] and

[1] The project is named after Florence Nightingale, who is seen as the founder of nursing sciences. When she worked as a nurse, she wandered through the hospital during the nights to look after her patients, why she became known also as the lady with the lamp.

D. Keyson et al. (Eds.): AmI 2011, LNCS 7040, pp. 309–313, 2011.

the evaluation of this system by seniors in a local care home. The knowledge that has been gained may be applied in the development of automatic gesture recognition systems that fit typical or natural human behavior and capabilities.

The main question which will be answered in this study is: What is the influence of interaction modality in the context of HRI on user acceptance and preferences? Simply said, when an elderly person performs a gesture/tactile command towards a robot screen, does that have influence on the users acceptance? And is there a preferred modality? The research question to some extent was inspired by the preliminary research regarding the acceptance of social robots by seniors [5], but predominantly they were chosen because of the importance to learn more about the perception by seniors of a social robot with multi modal interaction capabilities. The main research has been split in the following subquestions:

1. Does the HRI context afford a certain type of modality e.g. touch or gestures?
2. Which of the two modalities is preferred by the senior participants, or what are the objections for a particular modality against the other?
3. Is there a difference in gesture performance? Does the notion of $'Next'$ or $'Previous'$ lead to different gesture performances?

An experiment has been performed addressing these questions. In the experiment, participants were given the task to perform physical exercises to improve or maintain a healthy lifestyle. In order to move to the next exercise, the participants were asked to either *press a screen button which says next* (in case of the touch interface) or give a 'Next' In-Air gesture. No information was provided *a priori* about how to perform such a $'Next'$ or $'Previous'$. Thus, insight was gathered into human gesture perception of the actions $'Next'$ and $'Previous'$.

2 Design

For this comparative study between interaction modalities, a simple prototype of an assistive robot and an application have been developed. Both the application and the robot will be described in more detail here.

For this application a scenario has been chosen in which the elderly person performs exercises in order to improve lifestyle behavior. The senior in this scenario stands in front of the assistive robot. On the screen of the robot several body postures are presented that have to be copied by the senior. After each successfully performed posture (as recognized by the recognition part of the software) the senior navigates to the next or previous exercise.

HRI may be realized using different modalities such as speech, head pose, gesturing and touch or a combination of these modalities. This study compares two modalities namely touch and gestures. The main concern regarding the design and implementation of the software application was gesture recognition.

Gesture recognition is a very popular research area ([3,2])in which the implementation of various kinds of feature extraction algorithms finally result in the recognition of the points of interest such as a human hand.

Instead of traditional gesture recognition software, a contemporary approach is used in this research: The Microsoft Kinect 3D sensor array. (see figure 1a).

This 3D sensor was originally designed for the game console Xbox 360. But the 3D sensor is also usable when connected with the PC. The Kinect is mounted on a stand. The stand is mounted on top of the mobile platform PeekeeII [4]. A touchscreen, which essentially is a touchscreen enabled laptop is mounted below the Kinect(See figure 1b). By having the depth-of-field camera and the RGB camera a calculated distance apart, the Kinect is able to perform immediate, 3D incorporation of real objects into on-screen images.

(a) Kinect setup and interface **(b)** Robot

Fig. 1. Setup

3 Methodology of the Experiment

3.1 Subjects

Participants in the experiment were 12 elderly people who participated voluntarily in this study, and signed a consent form. The average age of the participants was $77,17(\sigma = 7.19)$ with the youngest being 71 and the oldest 96. Of the 12 participants 7 were female. 8 participants had mobility problems. Most of participants reported to never have used a computer before. The most frequent appliances used by the participants were the TV, coffee machine and microwave.

3.2 Experimental Setup

The gesture recognition system is implemented in such a way that it even if the 'Next' or 'Previous' gesture is not performed precisely as suggested it can be recognized correctly. Not only differences between modalities can be measured. Agreements in the way that gestures are performed may become visible as well as the different notions which the participants have towards the notion of *gestures* in general. Each participant was asked whether he or she knows what gestures are, and how he or she would perform a $'Next'$ or $'Previous'$ gesture before showing how the actual gesture should be performed in order for the system to be recognized.

3.3 Data Acquisition, Procedure and Analysis

The participant is recorded during the experiment. The Technology Acceptance Model(TAM) is used to investigate Effort, Ease & Anxiety (EEA) and Performance & Attitude (PA) [1]. Together with a short interview which is recorded on video, insight in the preferences and acceptance of the interaction modalities is obtained. A within subject design is chosen to measure differences between the modalities gestures and touch. Counterbalancing of the two modalities is a applied to avoid order effects. The participant started with filling in a pre-test with questions regarding their daily use of appliances. A questionnaire including questions regarding a modality was filled in after each modality experiment. A final interview was held in which comparing questions were asked regarding preference, effort, ease and attitude.

4 Results

An item in the gesture questionnaire asked whether the participants found the gestures easy to perform. Using a 7 -point Likert scale 12 subjects answered with a mode of 7 (6 out of 12 answered with a 7) and an average of 6.4. The exact same result is discovered after the analysis of the question regarding the touch modality wherein the question was asked whether the participant found it easy to press the screen buttons. Testing these results with a Wilcoxon signed-rank test to a neutral result, yielded $Z = -.587$, $p = 0.557$. No significant differences were found on the other questions of the questionnaire either.

The interview yielded valuable information concerning alternative gestures by the participants for the concepts 'Next' and 'Previous'. Also interesting behavior was noticed after the preliminary question about the notion of gestures. 4 out of 12 participants knew instantly what gestures and they even gave examples of gestures which they used back in the days during work or sports. Although they had different ideas about the performance of the gestures $'Next'$ and $'Previous'$, they did not have any problems understanding and relating the specified gestures to the concepts $'Next'$ and $'Previous'$.

In the final interview participants reacted more positive towards the use of in-air gestures. 9 out of 12 participants preferred the gesture interface. They also reported that they have little knowledge about assistive robots. They were inquisitive and felt the need to have more information which is expected to result in a overall higher level of robot acceptance. Many participants (7 out of 12) argued that they could express themselves more using in-air gestures as opposed to pressing screen buttons. Physical constraints of the participants was also a cause of the before mentioned preference, as they had to walk towards the robot in order to touch the screen.

5 Conclusion and Future Work

Two interaction modalities were compared; in-air gestures and touch. No significant results were found regarding the variables Effort, Ease & Anxiety (EEA) and Performance & Attitude (PA). The results on these variables were very high for both interaction modalities, indicating that both modalities were accepted by the elderly people.

The results on questions in the final interview where people were asked to compare the use of the two modalities indicate that the participants reacted more positive towards the use of in-air gestures. Most participants had a preference to use in-air gestures for the interaction with the robot because they could express themselves more using gestures as opposed to pressing touch screen buttons. An extra reason to prefer gestures were the physical constraints of many of the participants. In the touch interface they had to walk towards the robot in order to touch the screen. In-air gestures can be further applied in for instance *calling* the robot, as well as *interrupting* the robots activity.

Acknowledgements. This research is supported by the Florence project. Florence is supported by the European Commission in the FP7 programme under contract ICT-2009-248730.

References

1. Heerink, M., Kröse, B., Wielinga, B., Evers, V.: Measuring the influence of social abilities on acceptance of an interface robot and a screen agent by elderly users. In: Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology, BCS-HCI 2009, Swinton, UK, pp. 430–439. British Computer Society (2009)
2. Keskin, C., Erkan, A., Akarun, L.: Real time hand tracking and 3d gesture recognition for interactive interfaces using hmm. In: Joint International Conference ICANN/ICONIP. Springer, Heidelberg (2003)
3. Park, C.-B., Lee, S.-W.: Real-time 3d pointing gesture recognition for mobile robots with cascade hmm and particle filter. Image and Vision Computing 29(1), 51–63 (2011)
4. Wany-Robotics. PekeeII Essential package (2011), `http://www.wanyrobotics.com`
5. Znagui-Hassani, A.: Discovering the level of robot acceptance of seniors using scenarios based on assistive technologies. Technical report, University of Twente (HMI) (2010)