1-1-2016

Design and Development of the eBear: A Socially Assistive Robot for Elderly People with Depression

Amirhossein Kargarbideh
University of Denver

Design and Development of the eBear: A Socially Assistive Robot for Elderly People with Depression

Abstract

There has been tremendous progress in the field of robotics in the past decade and especially developing humanoid robots with social abilities that can assist human at a socio-emotional level. The objective of this thesis is to develop and study a perceptive and expressive animal-like robot equipped with artificial intelligence in assisting the elderly people with depression. We investigated how social robots can become companions of elderly individuals with depression and improve their mood and increase their happiness and well-being. The robotic platform built in this thesis is a bear-like robot called the eBear. The eBear can show facial expression and head gesture, can understand user's emotion using audio-video sensory inputs and machine learning, can speak and show relatively accurate visual speech, and make dialog with users. the eBear can respond to their questions by querying the Internet, and even encourage them to physically be more active and even perform simple physical exercises. Besides building the robot, the eBear was used in running a pilot study in which seven elderly people with mild to severe depression interacted with the eBear for about 45 minutes three times a week over one month. The results of the study show that interacting with the eBear can increase happiness and mood of these human users as measured by Face Scale, and Geriatric Depression Scale (GDS) score systems. In addition, using Almere Model, it was concluded that the acceptance of the social agent increased over the study period. Videos of the users interaction with the eBear was analyzed and eye gaze, and facial expressions were manually annotated to better understand the behavior changes of users with the eBear. Results of these analyses as well as the exit surveys completed by the users at the end of the study demonstrate that a social robot such as the eBear can be an effective companion for the elderly people and can be a new approach for depression treatment.

Document Type

Thesis

Degree Name

M.S.

Department

Computer Engineering

First Advisor

Mohammad Mahoor, Ph.D.

Second Advisor

Timothy Sweeny

Third Advisor

Kimon Valavanis

Keywords

Assistive technologies, Robotics, Socially assistive robotics, Social robots

Subject Categories

Engineering | Robotics

Design and Development of the eBear: A Socially Assistive Robot for Elderly People with Depression

A Thesis

Presented to

The Faculty of the Daniel Felix Ritchie School of Engineering and Computer Science

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Amirhossein Kargarbideh

March 2016

Advisor: Dr. Mohammad Mahoor

Author: Amirhossein Kargarbideh
Title: Design and Development of the eBear: A Socially Assistive Robot for Elderly People with Depression
Advisor: Dr. Mohammad Mahoor
Degree Date: March 2016

Abstract

There has been tremendous progress in the field of robotics in the past decade and especially developing humanoid robots with social abilities that can assist human at a socio-emotional level. The objective of this thesis is to develop and study a perceptive and expressive animal-like robot equipped with artificial intelligence in assisting the elderly people with depression. We investigated how social robots can become companions of elderly individuals with depression and improve their mood and increase their happiness and well-being. The robotic platform built in this thesis is a bear-like robot called the eBear. The eBear can show facial expression and head gesture, can understand users emotion using audio-video sensory inputs and machine learning, can speak and show relatively accurate visual speech, and make dialog with users. the eBear can respond to their questions by querying the Internet, and even encourage them to physically be more active and even perform simple physical exercises. Besides building the robot, the eBear was used in running a pilot study in which seven elderly people with mild to severe depression interacted with the eBear for about 45 minutes three times a week over one month. The results of the study show that interacting with the eBear can increase happiness and mood of these human users as measured by Face Scale, and Geriatric Depression Scale (GDS) score systems. In addition, using Almere Model, it was concluded that the acceptance of the social agent increased over the study period. Videos of the users interaction with the eBear was analyzed and eye gaze, and facial expressions were manually annotated to better understand the behavior changes of users with the eBear. Results of these analyses as well

as the exit surveys completed by the users at the end of the study demonstrate that a social robot such as the eBear can be an effective companion for the elderly people and can be a new approach for depression treatment.

Acknowledgements

I would like to express my deepest gratitude to my advisor Dr. Mohammad Mahoor for his support throughout this thesis, for his patience, and for his immense knowledge.

I would also like to thank the Eaton Senior Communities, in Particular Diana Delgado and Sarah Schoeder, for letting us running the pilot study in their center. Without their help and the seven residents who participated in the project, this study could not have been successfully conducted.

This work is partially supported by the National Science Foundation grant IIS-1111568.

Finally I would like to express my sincere appreciation to my parents, who always supported and encouraged me during my years of this study. Thank you.

Amirhossein Kargarbideh

Contents

List of Figures

List of Tables

1

Introduction

1.1 Introduction and Motivation

With the increase in the population of the aged people in the United States and all other countries in the world, there is a crucial need for a better personalized health care system. Socially Assistive Robotics (SAR), as part of assistive technologies, aims at providing healthcare for people and particularly for the aging population and decreasing the current healthcare services' costs. The estimated five-fold increase in the population of people over the age of 85 by the end of 2050 (9) points out the importance of such technologies. The ultimate goal is to treat abnormal social behaviors caused by stroke, childhood diseases, or depression as well as improving social skills in general. This study mainly focuses on dealing with elderly people, particularly those who suffer from depression.

Detection and prevalence of major depression has increased over the past 50 years (52). The World Health Organization has anticipated that depression will be the second cause of death and disability by the year 2020 which is a severe threat, specifically to adults. Depression can be recognized in the general population by various signs including increase/decrease in sleep, increase/decrease in appetite, anhedonia, poor concentration, and suicidality (21). Lack of mobility, which is one of the focuses of this thesis, is another symptom of depression. For instance, individual suffering from depression may spend hours sitting without much social interaction. In this situation, the best way is to increase the patients mobility such as asking them to complete some physically related tasks. The connection between exercise and depression is not completely clear but it has been proven that in the case of mild to moderate depression, regular exercise can change the individuals' mood (23). For example, in (60) they have several groups of elderly aging from 60 to 96 years and have concluded that the group that was continuously active had lower depression than the other groups. Hence, developing an automatic agent such as a robot capable of motivating the elderly to exercise throughout a day is of great interest. Here we discuss the taxonomy of social robots:

1.1.1 Social Robotics

The focus of this research is on the applications of social robotics in elderly care, another category of social robots which deals with kids is out of the scope of this research. There are different viewpoints toward social robot taxonomy (37) (36) (49). Each one targeted the needs of human users and applications of the robot from various perspectives. However, there is no straight line between different categories. In most robotic platforms, usually there is more than one type involved. Feil-Seifer (35) categorized the social robots into three different categories: Assistive Robotics, Socially Interactive Robotics, Socially Assistive Robotics. Here we describe each briefly:

1.1.1.1 Assistive Robotics

Assistive robots (AR) used to be defined as the type of robots which are used to help people in a physical manner such as assistive manipulator hands or rehabilitation robots. Even though this definition includes a vast majority of such robots, it does not cover the robots that aid people in a non-physical manner like the ones assisting elderly people with socio-emotional disorders. Therefore, a comprehensive definition for Assistive Robots is the robots that assist or support a person (35). The type of assistance differs based on the use and place, i.e. care centers, homes, schools.

1.1.1.2 Socially Interactive Robotics

Socially interactive robots (SIR) are the ones in which social interaction plays a key role. According to Fongs description (37) of socially interactive robots, they should exhibit such features: express and/or perceive emotions; communicate with high-level dialogue; learn/recognize models of other agents; establish/maintain social relationships; use natural cues; exhibit distinctive personality and character; may learn/develop social competencies. Depending on the purpose and the application, social robots can have different shapes and functions. Some of the socially interactive robots engage users in conversations pro-actively, while other ones wait for the prompt from the user to start the conversation.

1.1.1.3 Socially Assistive Robotics

Unlike assistive and socially interactive robotics, socially assistive robotics (SAR) aid people through non-physical social interaction. Feil-Seife (35) defines socially assistive robotics as the intersection of socially interactive robotics and assistive robotics. The goal of assistive robotics is to assist people, whereas the goal of socially interactive robots is to interact with human users for the sake of interaction. In socially assistive robotics, it aims at interacting with people in a social manner in order to assist and support them in different areas such as depression, dementia, rehabilitation.

SAR for the elderly individuals is a relatively new area and there are a limited number of works with this focus. In (63) a non-autonomous robot was designed as an exercise demonstrator but it does not have any sensors and therefore no feedback is provided. In (25) and (55) they have developed agents with the aim of assisting the elderly people in exercising, however they do not monitor the way the user is performing the exercise. Also their system is mostly conversational rather than interactive. The other SAR example is the Social exergame (26) (48) (94) which includes a broad range of games that require the user to use a remote controller or a motion sensor to play a game. These games aim at persuading seniors to increase their physical activity. However, as (87) points out, these games may not be the perfect match for the elderly, because the games are usually so fast or provide negative feedback sometimes. Also, there is a risk of falling in such games. One motivation for the SAR is the fact that not always physical contact is the best type of help for a human user, in some cases the user needs to be treated in a social manner. A good example is the post-stroke patients whom need a therapist to remind them of the limbs which need to be exercised. This type of rehabilitation is called Constraint Induced movement therapy (93). Furthermore, since SAR does not include physical contact with the patient, it reduces the risk of falling or any other related damages. There is a type of treatment called pet-therapy. Some elderly people create closer and more effective interaction with pets rather than human therapists. Following this inclination, using robots to assist the elderly might have better results. The social robot developed in this thesis combines the features of various social robotic research as well as several unique features which are discussed in the next chapters.

1.2 Contribution

There has been tremendous progress in the field of robotics in the past decade and especially developing humanoid robots with social abilities that can assist human at a socio-emotional level. The objective of this thesis is to develop and study a perceptive and expressive animal-like robot equipped with artificial intelligence in assisting the elderly people with depression. We investigated how social robots can become companions of elderly individuals with depression and improve their mood and increase their happiness and well-being. The robotic platform built in this thesis is a bear-like robot called the eBear. The eBear can show facial expression and head gesture, can understand users emotion using audio-video sensory inputs and machine learning, can speak and show relatively accurate visual speech, and make dialog with users. the eBear can respond to their questions by querying the Internet, and even encourage them to physically be more active and even perform simple physical exercises. ”Autonomous” is the term which made the eBear unique. The eBear successfully interacted with 7 elderly people with depression for a month in an automatic manner. To best of our knowledge, the eBear is the first social robot containing all the following features in one platform:

- The eBear is semi-autonomous in the sense that all the modules are run automatically, but the order and the time at which the modules were run were determined by the operator.
- The eBear could be used in a home or care-center setting without the need for any special setup.
- The eBear could analyze the emotions of users from both visual and speech cues at the same time.
- The eBear could automatically run a mood evaluation method which could be used to assess the mood remotely.
- The eBear is proactive.
- The eBear could engage and carry out an open conversation.
- The eBear could motivate and couch users through exercise and provide them with the required feedbacks.

Besides building the robot, the eBear was used in running a pilot study in which seven elderly people with mild to severe depression interacted with the eBear for about 45 minutes three times a week over one month. The results of the study show that interacting with the eBear can increase happiness and mood of these human users as measured by Face Scale, and Geriatric Depression Scale (GDS) score systems. In addition, using Almere Model, it was concluded that the acceptance of the social agent increased over the study period. Videos of the users interaction with the eBear was analyzed and eye gaze, and facial expressions were manually annotated to better understand the behavior changes of users with the eBear. Results of these analyses as well as the exit surveys completed by the users at the end of the study demonstrate that a social robot such as the eBear can be an effective companion for the elderly people and can be a new approach for depression treatment.

2

Related Works

Socially Assistive Robotics (SAR) for the elderly people is a relatively new area and there are a limited number of works with this focus. In (63) a non-autonomous robot was designed as an exercise demonstrator but it does not have any sensors and therefore no feedback is provided. In (25) and (55) they have developed agents with the aim of assisting the elderly in exercising, however they do not monitor the way the user is performing the exercise. Also their system is mostly conversational rather than interactive. Other SAR example is the Social exergame (26) (48) (94) which includes a broad range of games that require the user to use a remote controller or a motion sensor to play a game. These games aim at persuading seniors to increase their physical activity. However, as (87) points out, these games may not be the perfect match for the elderly, because the games are usually so fast or provide negative feedback sometimes. Also, there is a risk of falling in such games.

(33) is another research which was covered in the eBear. They have developed a robot exercise instructor to monitor and motivate the elderly to perform physical activities. Their robot is capable of three exercise games with the user. However, they require the user to sit in front of the robot with a dark curtain behind him/her. Also, the user communicates

with the robot via a Wiimote remote control. They have concluded that their SAR system is intelligent and helpful and gives the user a good mood.

Researchers have been working in the area of social robotics since around two decades ago. In order to have more organized literature reviews, we continue with introducing some of the well known robots and then a few of the related research is reviewed in each case. Table 2.1 illustrates the main limitations of the similar works which are addressed in the eBear.

2.1 Paro

Paro is a seal-like interactive robot developed by the National Institute of Advanced Industrial Science and Technology in Japan (18). Paro has been under development since 2003 and as of the time of publication of this research, the 8th generation is out (Figure 2.1). It has few actuators to respond to the outside environment as well as imitating the voice of a real baby seal. In addition, it is equipped with five types of sensors: light sensor to recognize the brightness of the environment (dark or light), tactile sensor to feel when the robot is being held or stroked by the user, Audio sensor for speech recognition and detecting the direction of the sound, temperature sensor, and posture sensor. Guinness World Record has certified this robot as the most therapeutic robot in the world (18). Because of its appearance and functionalities, it has been shown that it is capable of reducing the stress of the patient as well as improving the interaction and communication between the patient and the caregiver (18).

(72) has evaluated Paro robot in a multi-sensory behavioral therapy. Since Paro was used in a group of adults with cognitive impairments, they concluded that both the people who were interacting with Paro and the people who were just in the environment have had improvements in their activity levels. In (90) and (51), they used Paro in a care house

Table 2.1: Main limitations of the similar works which are addressed in the eBear

Robot	Speech Recognition	Open Conversation	Setup Not Required	Proactive Engagement	Coach Exercise	Various Range of Functionalities (i.e., games)	Automati Mood Evaluation	Facial Expression Generation	Facial Emotion Recognition	Speech Emotion Recognition
Paro	No	No	Yes	No	No	No	No	No	No	No
AIBO	Yes	No	Yes	Some	No	No	No	Some	Yes	No
Bandit	No	No	No	No	Yes	No	No	Yes	No	No
PaPero	Yes	No	Yes	Some	No	Some	No	No	No	No
NeCoRo	Yes	No	Yes	No	No	No	No	No	No	No
iCat	Yes	No	Yes	Some	No	No	No	Yes	Yes	No
Healthbot	Yes	No	No	Some	No	Yes	No	No	No	No
The eBear	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

Figure 2.1: Paro Robot - A seal-like interactive robot

and let the elderly residents to freely interact with the robot up to 9 hours per day. They concluded that Paro encouraged the residents to communicate better with each other. They also showed that there were improvements in the reactions of the subjects vital organs to stress. (91) reports the results of an on-going 5 year experiment on using Paro for elderly people. They showed that Paro improved people mood and the interesting result is that this improvement showed up during five years and the relationship between the elderly people and the robot continued during that period.

2.2 AIBO

AIBO or Artificial Intelligence Robot means ”companion” in Japanese and is a dog-like robot developed by Sony (20). As it can be seen in Figure 2.2, it is equipped with a wide range of sensors including touch, camera, range-finder, microphone, acceleration, and angular velocity. AIBO is able to shape its personality based on the interactions with its owner and surroundings. It has come in three generations and was discontinued by the manufacture in 2006.

In (41) they have developed two recreation games using AIBO for people with dementia. In their games they targeted the ability of memorization and the ability of emotion

Figure 2.2: AIBO - A dog-like robot developed by Sony

control and accommodation to society. One of the games was card game which was tested on a one-to-one basis and the other game was a ball game which was done in group. By 5 days of deploying the robot in the field (each game was done once a day), both the care-givers' evaluation at the nursing home and the evaluation that was done by the game showed ability improvements in their subjects. (22) uses AIBO to compare a pet-like robot to a real dog when it comes to treating loneliness. They had 13 control group and let the residents of the care facility to interact with both the real dog and AIBO 30 minutes per week for each. They showed that both the pet-like robot and the real dog can reduce loneliness. (71) conducted a study on acceptability of AIBO, or an artificial companion in general, by cognitively impaired elderly. They showed that negative feelings towards technology do not affect the interaction. However, to achieve awareness of the usefulness of the companionship, positive attitude is required. For several other research on AIBO, the reader is referred to (80), (73), and (64).

2.3 Bandit

Bandit is one of the robots used in the University of Southern California's Robotics Research Lab (7). It has come in two generations and the current generation has 22 degrees of freedom in the arms, head, neck, waist, eyebrow, and mouth. Bandit (Figure 2.3) is a humanoid upper-torso robot which is mounted on a ActivMedia Pioneer 2DX mobile robot equipped with a range finder, speaker and a Sony Pan-Tilt-Zoom camera.

Figure 2.3: Bandit - A humanoid upper-torso robot of University of Southern California

In (81) they performed a study using Bendit on people with dementia and/or cognitive impairments. They used Bandit to improve or at least maintain the participants' cognitive attention by encouraging a music-based cognitive game. Their pilot study consisted of 9 residents of a senior living care facility. Each participant interacted with the robot 20 minutes per week for a duration of 6 months excluding the 2 months of learning. They concluded that people with dementia and/or Alzheimer can maintain their attention to music during a long period of time. (86) used Bandit to show the effectiveness of embodiment in compare to a screen agent. Similar to their other work (81), they used Bandit to encourage a music-based cognitive game. They performed a post-experiment evaluation using the Standardized Mini-Mental State Examination (SMMSE) cognitive test. The participants' scores were either maintained or improved after interaction with Bandit for 8 months. In (33) they used Bandit as an exercise coach. The robot would monitor the user during an exercise and the aim was to keep the user motivated to finish the exercise. It also would provide some feedbacks on how to improve the exercise. They need the user to sit on a chair with a curtain behind them. There were no interactive conversations since the user had to respond via a remote controller with a few buttons on it. Apart from the exercise, user had the minimal interaction with the robot which does not have much overlap with the definition of SAR according to (35). For several other studies on Bandit, the reader is referred to (82), (84), (85), and (83).

2.4 PaPero

PaPero stands for Partner-Type Personal Robot which is developed by Japanese firm NEC. It is a personal robot with a human-like baby face (figure 2.4). It has been designed and developed with the purpose of becoming human's partner so that they can live together. This 6.5kg robot is capable of recognizing speech, synthesizing speech, recognizing faces, and reacting to human's touch (3). PaPero is also equipped with cameras and ultrasonic sensors so that it can walk around the place without colliding other objects. The good part about PaPero is that it comes with its own development environment. Using the environment, the developer can easily add different actions, behaviors, and conversations to the robot.

PaPero has been used as a socially assistive robot in Australia with the name of Matilda. In (54) they describe the embodiment of multimodal interaction modules such as gesture,

Figure 2.4: PaPero - A personal robot developed by Japanese firm NEC

voice, emotion etc in Matilda. They targeted both one-to-one and group based interactions. The study was done on 34 elderly participants aging from 71 to 98 having various symptoms such as dementia, Parkinson disease, depression, and memory loss. The robot was deployed in three nursing houses and participants interacted with the robot for four days. This experiment, which is claimed to be the first field trial in Australia, was evaluated using different measures including questionnaires, observation, recognized facial emotions of the robot etc. They concluded that such a robot would improve the wellbeing of the participants as well as care personalization. This study was reported in another publication as well (53) .

2.5 NeCoRo

NeCoRo is a commercial robotic cat developed by Omron Corporation in Japan. It has 15 actuators so that it can responds to others behaviors. It is also covered with synthetic fur which feels natural and looks like a real cat (Figure 2.5). This 1.6 Kg robot is equipped with different touch sensors all around its body to react to the user's touches. One interesting point about NeCoRo is that its personality gets adjusted to its owner. For example, if the user call a name over and over, then it will remember its name and will react when that name is called. Other included sensors are sight, sound, and orientation. NeCoRo provides both verbal, such as mewing, and nonverbal, such as tail wagging responses.

Figure 2.5: NeCoRo - A catlike robot developed by Omron Corporation

In (66) they did a study on the relationships between robots and human, and the type of effects the robot might have on the user. There were 12 males and 21 females. They interacted with NeCoRo for a period of one year in a Elderly-care facility in Japan. Through their studies, they concluded that interacting with NeCoRo made the users' expressions livelier by having them touching the robot, talking to it and picking it up. It also resulted in making the user more comfortable and happy via better communications with the robot. These results were gathered based on staff comments as well as questionnaires which were filled by staff members. There is another study on using NeCoRo to evaluate the effects of an interactive robot from psychological point of view (59). Among 32 people who participated in their experiment, there were 16 Americans and 16 Japanese. These people were from two age groups of 20-35 and 65-79. They only intrectaed with the robot for 15 minutes. For evaluation purposes, they designed a method called Person-Robot Complex Interactive Scale (PRCIS). This evaluation method consists of different criteria including biological, psychological, and social factors. The interesting part of their study is that they evaluated the robot on people with different diversities (age, geographical). They showed that older people like the interactions better than the younger age group. They also concluded that males enjoyed the robot's active behavior more than females.

2.6 iCat

iCat is a research platform developed by Philips Electronics (19) with the goal of stimulating human-robot interaction research. This 38 cm tall cat-face robot is equipped with 13 servos which control the face, eyes, eyebrows, eyelids, head, and mouth (Figure 2.6). These servos make the robot capable of making different facial expressions such as sad, happy, angry etc. A speaker, microphone, camera, and touch sensor are also installed on iCat.

In (42) they used iCat to examine how users would perceive such a robot with two different social capabilities. For this purpose, they designed two different interaction scenarios where in one of them the robot was more socially interactive. More socially interactive means it looks at the user and nods the head while the user is speaking. It also remembers

Figure 2.6: iCat - A research platform developed by Philips Electronics (12)

and uses the user's name while speaking as well as showing different facial expressions. However, the dialog was the same for both scenarios and the difference was not in the conversations. 40 elderly people participated in their study which took place in Netherlands. iCat was used in a Wizard of Oz manner to make sure the interaction pattern was the same for all the participants. First they met iCat in a 5-minute group section where the robot told them about its features and the ones that will be done in the private section. Then participants had a conversation with the robot in which three simple tasks was done by iCat. The study was evaluated using post-interaction questionnaires. They assessed the robot from function and conversation point of view. They did not find any significant difference in using the two modes in terms of robot acceptance. However, participants were more comfortable with the more sociable version as a conversational partner. The same group conducted another study in (43) to evaluate the effect of iCat versus a screen agent in two modes of less and more expressive conditions. The screen agent was a human-like character which was capable of expressing different facial expressions just like iCat. The same as the other study 40 elderly people had a group session as an introduction. Then iCat performed some simple functions such as information providing, agenda keeping etc. The screen agent performed other tasks such as alarm setting and giving direction which are not the same as iCat. They concluded that people showed more expressiveness to the more expressive versions of iCat and the screen agent and this effect is stronger in the case of iCat. However, these results are difficult to be adjusted since the robots performed different tasks and they had different appearances. (44) is another study by the same group on iCat which is done in a similar fashion of comparing a social robot with a screen agent. They had three hypothesizes and they expected that social presence would result in enjoyment, enjoyment would result in intention to use, and intention to use would result in the user actually use the product. These hypothesizes were confirmed. As they mentioned, in these studies, people interacted with iCat for only 10 minutes. Hence, the results are probably affected by the fact that the system was new and interesting for the user and in a long run, with high probability, it might become boring.

2.7 Healthbot

Healthbot is a joint project between Yujin Robot, University of Auckland and Electronics and Telecommunications Research Institute (ETRI) in Korea. Unlike the aforementioned robots, Healthbot does not have a human-like or animal-like appearance and it is mostly used as a service robot(figure 2.7). It is equipped with a battery-powered differential drive. Localization is done using a StarGazer system (16) which requires the ceiling of room to have passive landmarks. This requirement makes the robot not convenient to be used for daily life purposes. Healthbot is also equipped with a microphone, bumper sensor, ultrasonic sensor, and a touch screen.

Figure 2.7: Healthbot - Healthbot is a joint project between Yujin Robot, University of Auckland and Electronics and Telecommunications Research Institute (ETRI) in Korea

(78) used Healthbot to evaluate if emotions and attitudes towards the robot are predictors of how they accept the robot. In their study 32 residents and 21 staff of a retirement village in New Zealand interacted with the robot for a period of 30 minutes. The robot was designed with seven service applications such as greeting, medication reminding, vital signs measurement etc. They were given two questionnaires, one before the interaction and one after the interaction in order to evaluate the effects of interacting with the robot. The questionnaire included two scales, Robot Attitude Scale (RAS) and Positive and Negative Affect Schedule (PANAS). Decrease in negative affect towards the robot as well as more favorable attitude were resulted after interaction with the robot. They also concluded that changes in emotions and attitudes are predictors of the robot acceptance.

In (47) they used the outcomes of (78) to introduce the second version of Healthbot for elderly people. In this version they put more focus on the user's need so that the robot can be customized in the field. They conducted three parallel studies at a retirement center in New Zealand. One study was putting the robot in the common areas of the center so that people can have free interaction with the robot for two weeks. The other one was the same but in private rooms for two weeks, and the last study was to monitor falls remotely. In case of a fall the robot would go to the place and perform monitoring session. 67 people with age 65 or more participated in this study. They evaluated the robot with three questions including how they enjoyed the robot, how they rate the interaction, and how they would like to interact with the robot again. Although these questions are good to evaluate the overall effectiveness of the robot but the questions are rather general to specify the actual opinion of people and how the robot affected them. However, as they mention, this was an ongoing research.

3

Hardware and Software Design of the eBear

3.1 Hardware Design

This chapter explains the mechanical design, development, and the features of the eBear. Expressive Bear or the eBear is a bear-like robotic platform developed at the University of Denver Computer Vision Laboratory. We introduced the first version of the platform in (96). Figure 3.1 depicts the first version of the eBear. Basically, the eBear is a mechanical face equipped with servos which is mounted on a pedestal box that is covered with fur (Figure 3.2).

3.1.1 The eBear Platform

In this section the mechanical platform of the eBear is described. First the initial version of the eBear is explained. After that, the modifications which were made to the platform are discussed. the eBear is equipped with 10 degrees of freedom (DOF) in its head and the rest of its body is still. As it can be seen from figure 3.3, the degrees of freedom are

Figure 3.1: eBear - Initial version of the eBear Platform

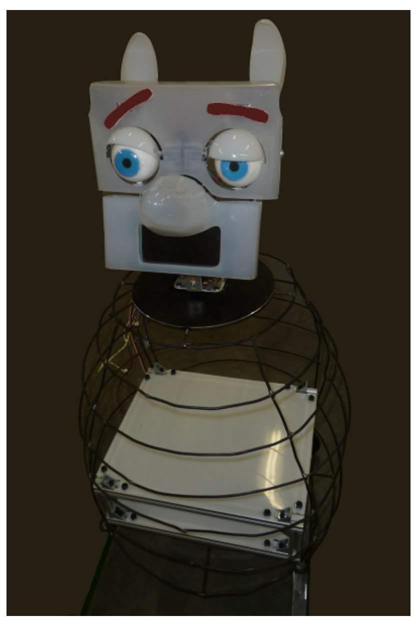


Figure 3.2: Internal Structure of the eBear - The eBear is a mechanical head mounted on a pedestal box

left eyebrow (f1), right eyebrow (f2), forehead tilt (f3), eyeballs (f4), left eyelid (f5), right eyelid (f6), left ear (f7), right ear (f8)), neck pitch and yaw (f9, f10).

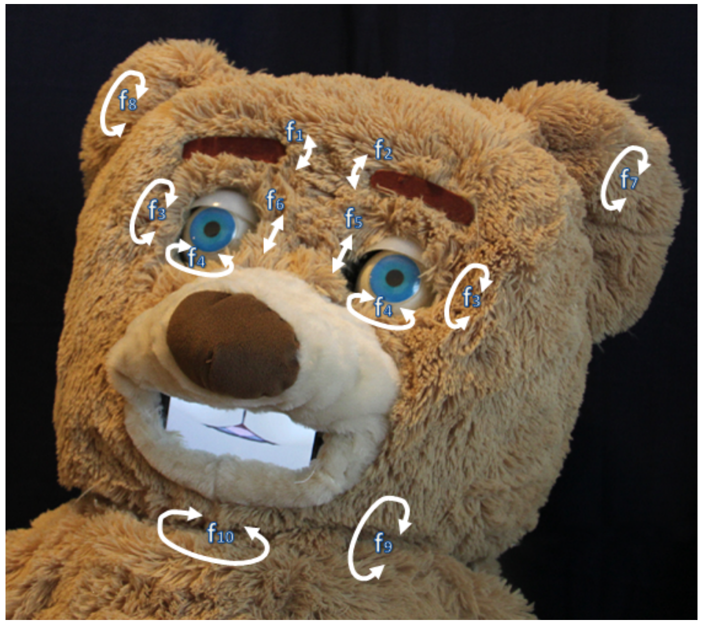
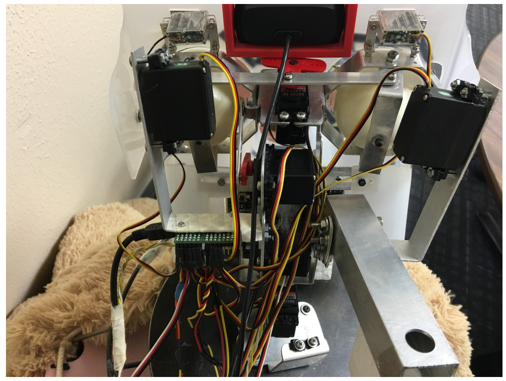


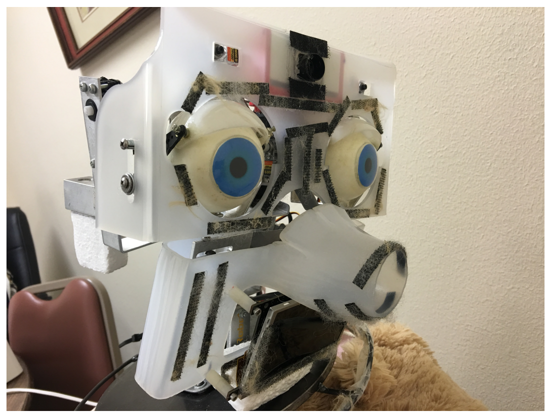
Figure 3.3: The eBear degrees of freedom - 10 degrees of freedom of the eBear (96)

Figure 3.4 illustrates the internal structure of the head. Each degree of freedom is controlled with a "Hitec" PWM servomotor (2). "Mini Maestro (10)" servomotor controller is used to control the servos. The commands are received from a C# .Net Framework software.

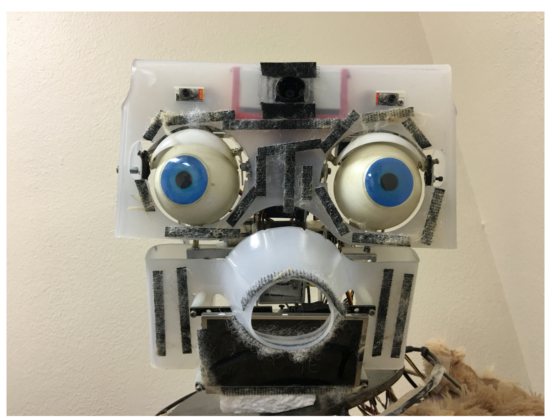
One of the features that makes the eBear unique is the hybrid face. The problem with mechanical mouth designs is the lack of enough flexibility for expressing various lip motions. The eBear's head is equipped with 10 degrees of freedom but the mouth is replaced with a LCD display, a 4.3" TFT LCD panel by Sharp. The resolution of the screen is 480×272 pixel in 16 bit colors. The LCD can be programmed with OpenGl using the provided APIs. An artist designed the models of visemes in Maya Software. By Blending proper models with different weights, a natural visual speech was obtained. Bavieca speech aligner (13) is used to time-align phonetic transcription of the recorded utterances. The an-



(a)



(b)



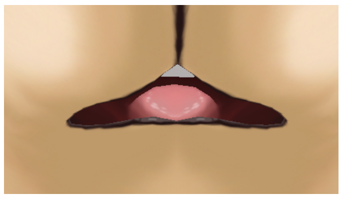
(c)

Figure 3.4: The internal structure of the eBear's head

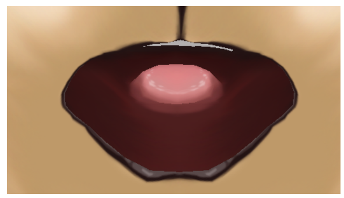
(a) Anger

(b) Joy

(c) Sadness

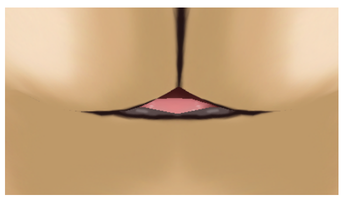


(d) Disgust



(e) Surprise

(f) Fear



(g) Neutral

(h) Viseme ‘O’

(i) Visime ‘m’

Figure 3.5: Examples of some visemes and expressions

imation system receives the time-aligned phonetic input from Bavieca speech aligner and produces the related visemes. The visemes specifies the movements of the lips and tongue which are alignes with the recorded speech. In order to better express the emotions of the eBear, the lip movements and the visemes are blended:

$$F_j = F_c + \lambda_j (F_j^{max} - F_0)$$

where F_c is the current viseme, F_j^{max} is the desired expression at the maximum intensity, F_0 is the neutral mode, and $\lambda_j \in [0,1]$ is the intensity of F_j. Figure 3.5 illustrates a few of the expressions which are shown on the LCD display.

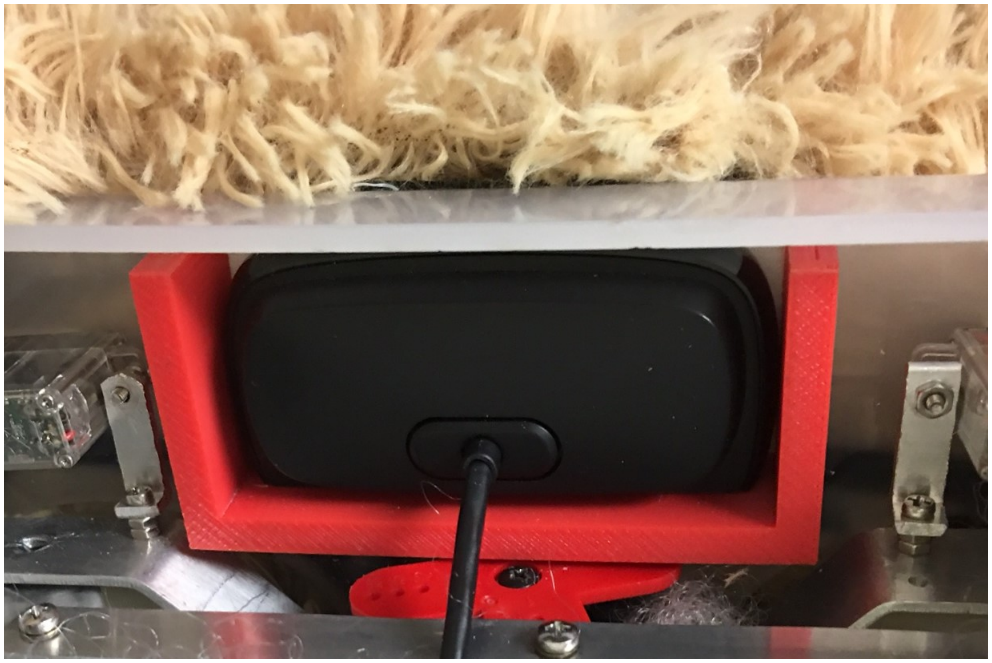
3.1.1.1 Camera

The platform shown in Figure 3.1 was the initial version of the eBear. However, for the purpose of this study, the eBear's body were modified and several hardwares were integrated to it. In the first version, a camera was mounted on the top of the eBear's head. The camera was not secure enough and it could move easily. In order to be able to use the camera for face tracking as well as facial expression recognition, it had to be mounted in such a way to minimize any kind of movement with respect to the head. A 3D bracket was designed in SolidWorks software and printed using a 3D printer so that the camera could fit inside that (Figure 3.6a). The bracket was glued to behind of the forehead plate and a hole was made on the head skin so that the camera could see the outside (Figure 3.6b). The camera is a HD Logitech webcam with resolution of 1280×720 pixels.

3.1.1.2 The Kinect Sensor

Microsoft Kinect sensor is equipped with three sensors which provide extensive information regarding the environment around as well as the people in front of it. We use a Kinect sensor on the body of the robot to monitor and capture the physical movements of the user. The Microsoft Kinect sensor version 2 is consisted of a depth sensor, an RGB camera and a microphone array (**?**). The depth sensor uses an infrared laser projector and an IR camera to capture depth images of 512×424 pixel resolution at 30 frames per second. The minimum depth distance is 50 centimeters and the maximum is around 4.5 meters. The horizontal field of view is 70 degrees and the vertical field of view is 60. It recognizes 26 joints of the body and can track up to 6 full skeletons. The color camera captures images of 1920×1080 resolution at 30 frames per second.

As shown in Figure 3.1, the first version of the eBear did not have any space to integrate the Kinect sensor. We took of the eBear skin and took the mechanical platform(Figure 3.2)



(a) The camera bracket

(b) The camera hole

Figure 3.6: The camera bracket integrated behind the forehead plate and the camera hole .

Figure 3.7: The Kinect Sensor - The second version of Microsoft Kinect sensor

to the machine shop of the University of Denver. The front part of one of the body wires were cut and two supporting wires were welded to secure the Kinect frame. A half semi-circle pad was designed in SolidWorks software to match the cross section of the eBear, since the cross section is not a perfect circle. The Kinect base frame was 3D printed and mounted inside the robot(Figure 3.8).

3.1.1.3 Screen

One of the purposes of the eBear was to engage the elderly people in activities such as exercise or games. For this purpose a screen was required to be integrated into the bear. Similar to the camera and the Kinect sensor, the eBear's platform was not designed to have such a screen. A frame was designed in SolidWorks software and was 3D printed. As it is shown in figure 3.10a, the frame is designed in a way to let the screen slides into it(Figure 3.10b). Another frame was fabricated in the University of Denver Machine Shop which was attached to the belly of the eBear. The 3D printed frame fits into the main frame in a way that only the screen is visible to the user (Figure 3.11). The screen is an Intel-based

(a)

(b)

Figure 3.8: An inside look to the Kinect base and how it is mounted inside the eBear.

Figure 3.9: The mounted Kinect sensor - An outside look to the mounted Kinect sensor

Quad Core HP tablet. A small window was cut out of the tablet frame so that the charger and USB cable could be plugged in through that.

3.1.2 The eBear's Facial Expression Generation

One of the challenges of developing the eBear was to design realistic facial expressions using the 10 degrees of freedom and the mouth animation. Facial Action Coding System (FACS) was introduced in (32) to taxonomize human facial expressions. In this method, each facial expression is expressed as a combination of different FACS units. FACS is designed for human faces and most of the units cannot be properly transfered to animals' faces. The eBear is an animal-like companionbot which is able to talk. Hence, we wanted the eBear to have both human-like and animal-like facial expressions. The eBear has 10 degrees of freedom which enables it to create a few of the FACS units. In order to create animal-like facial expressions, Darwin's interpretation of the expression of the emotions in animals (28) were used.

(a) **(b)**

Figure 3.10: A frame was 3D printed so that the screen slides into it.

Figure 3.11: The 3D printed frame is fit into the main frame

3.1.2.1 Design of Facial Expressions

(77) introduces a systematic methodology on how to design on facial expressions for robot with limited degrees of freedom. The design is based on psychological literature about emotional facial expressions and has been proved to be effective. According to their methodology, the first step is to project all the Action Units (AU) to the available degrees of freedom of the robot's face. The second column of table 3.1 illustrates the AUs related to each of the expressions of column one. Third column of Table 3.1 shows the degreees of freedom which are chosen for each specific expression. Since the face is hybrid, the mouth-related AUs are shown on the LCD. In case of mouth, there is no limit on portraying facial expressions. Because of the limited degrees of freedom on the rest of the head, a few of the AUs such as Cheek Raiser (AU6) or Brow Lowerer (AU4) could not be displayed. The AUs which the robot was able to make are shown in bold. One challenge of designing facial expressions for robots is the disgust emotion. The key part of disgust emotion is the nose wrinkle which is difficult to be shown on robots' face. The eBear has a forehead plate where the servos of eyebrow, eyelid, and eyes are placed in. This forehead plate is able to move up and down (f_3 in Figure 3.3).

As it was discussed, the third column of table 3.1 illustrates the projection of AUs to available degrees of freedom of the face. We wanted the facial expressions to be as expressive as possible. The eBear is a bear-like robot but the third column of table 3.1 does not utilize the animal side of the eBear. Darwin in (28) talks about the expression of emotions in human and animals. This study inspired us to use his definitions of animal emotions in the design of the eBear's facial expression. He describes animal's emotions in terms of various organs of their body, however we were only interested in head and face movement. Below are some of the observations that Darwin made:

- ”When mammals are suddenly frightened, as by a thunderstorm, or when they are made angry, as by being teased, their hair become erect. (P. 95)”
- ”The ears through their movements are highly expressive in many animals. (P. 110)”
- ”In very many animals, whenever they feel slightly savage, or pretend in their play to be savage, their ears are drawn back. (P. 111)”
- ”When a dog feels pleased and is caressed by his master, the ears fall down and are drawn back slightly. (P. 111) ”
- ”Some kinds of monkey, which have moveable ears, and which fight with their teeth, draw back their ears when irritated just like dogs. (P. 114)”
- ”All animals, when they are startled, or when they closely observe any object, direct their ears to the point towards which they are looking, in order to hear any sound from this quarter. (P 114)”
- ”The movements of a dog whilst exhibiting affection: these consist in the head and whole body being lowered and thrown into flexuous movements. The ears fall down are are drawn somewhat backwards, which causes the eyelids to be elongated. (P. 117)”
- ”Attention is shown by the head being raised, with ears erected, and eyes intently directed towards the object. (P. 121)”
- ”Dogs, when feeling affectionate, lower their ears in order to exclude all sounds. (P. 118)”
- ”A dog is disappointed with his head, ears, body, tail, and chops drooping, and eyes dull. (P. 120)”

- ”When dogs feel fear, ears are drawn backwards; but they are not pressed closely to the head, as in snarling, and they are not lowered, as when a dog is pleased or affectionate.”

Based on these observations it can be concluded that ears are highly expressive organ among many animals. As an example, when dogs are pleased their ears are drawn back and fall down, in case of fear ears are drawn back but without being close to the head, to show attention ears are erected. We also consulted with a few of dog owners and they were mostly agree with the descriptions of Darwin’s book. In addition, several cartoon animations such as ”Kung Fu Panda” and a few teddy bear animations were watched to have a better understanding of emotion expression among animals, in particular bears. These cues were taken into consideration while designing the animal-like facial expressions for the eBear. The fourth column of Table 3.1 illustrates the redesigned degrees of freedom involved in each of the expressions.

Table 3.1: Corresponding DOF in AU based and AU+Animal based expressions

	FACS AUs	**Corresponding DOF**	
		AU based	**AU+Animal based**
Happiness	AU6, **AU12**	LCD*	LCD, f_7, f_8
Sadness	**AU1**, AU4, **AU15**	LCD, f_1, f_2, f_5, f_6	LCD, f_1, f_2, f_3, f_5, f_6, f_7, f_8
Fear	**AU1**, **AU2**, AU4, **AU5**, **AU20**, **AU26**	LCD, f_1, f_2, f_5, f_6	L, f_1, f_2, f_3, f_5, f_6, f_7, f_8
Disgust	**AU9**, **AU15**, **AU16**	LCD, f_3	LCD, f_3, f_7, f_8
Anger	AU4, **AU5**, AU7, **AU23**	LCD, f_1, f_2, f_5, f_6	LCD, f_1, f_2, f_5, f_6, f_7, f_8
Surprise	**AU1**, **AU2**, **AU5B**, **AU26**	LCD, f_1, f_2, f_5, f_6	LCD, f_1, f_2, f_3, f_5, f_6, f_7, f_8, f_9

Figure 3.12 depicts all the two versions of emotions on the eBear’s face. Note that some parts of the expressions are dynamic and cannot be seen in the pictures. To express the joy of the eBear, ears are moved in reverse directions for 0.5 to 1.5 seconds based on the intensity of the emotion(3.12c). In sadness, ears are drawn forward along with eyelids are slightly closed. For fear, the forehead tilt is used to look downward while ears are drawn

(a) Joy: AU based

(b) Joy: AU+Animal based

(c) Sadness: AU based

(d) Sadness: AU+Animal based

(e) Fear: AU based

(f) Fear: AU+Animal based

(g) Disgust: AU based

(h) Disgust: AU+Animal based

(i) Anger: AU based

(j) Anger: AU+Animal based

(k) Surprise: AU based

(l) Surprise: AU+Animal based

Figure 3.12: Examples of expressions on the eBear.

back. Disgust is expressed with ears being drawn forward as well as the forehead tilts looks downward so that a kind of wrinkle is created around the nose. In anger, ears are erected to show the angriness. We called the simple expressions ”AU based” and the modified version ”AU+Animal based”.

We designed the bear to not only have facial expressions, but also be able to make expressions with different intensities. For this purpose, for each degree of freedom f_i there is a f_i^0 and a $f_{i,j}^{max}$ which denotes the natural state and the maximum expressiveness for that degree of freedom. Hence, the movement of each degree of freedom can be expressed as:

$$f_{i,j} = f_i^0 + \mu_{i,j}(f_{i,j}^{max} - f_i^0) \quad (3.1)$$

where, $\mu_{i,j} \in [0,1]$ represents the intensity of f_i.

3.1.2.2 Evaluation of Designed Facial Expressions

An experiment was designed to compare the effectiveness of AU based and AU+Animal based expressions. In this experiment, subjects interpreted the eBear's facial expressions. 21 subjects were chosen with the age ranging from 21 to 51 years old. They were selected from various cultural backgrounds and none of them were exposed to the eBear before the experiment. 6 AU based and 6 AU+Animal based expressions were shown to them in a random order. It took around 1.5 seconds for the robot to start with neutral emotion and change to the final position of one of the six expressions: Joy, Surprise, Sadness, Disgust, Anger, Fear. Participants were supposed to choose of the expressions on the questionnaire and they could take as much time as they needed. There was also a ”None” option in case they could not recognize the correct expression.

Table 3.2 and 3.3 illustrate the confusion matrices of the AU based and AU+Animal based expressions respectively. The recognition rate in case of Joy, Anger, Disgust, and Surprise were increased in AU+Animal based mode in compare to AU based mode. In AU+Animal mode, subjects confused Sadness with Disgust more than AU based mode which resulted in a 4.8% decrease in the recognition rate. One possible for this confusion is that in the AU+Animal mode of Sadness, the forehead tilt is involved which is common with Disgust. Fear recognition rate was also decreased from 33.3% in AU based to 19.0% in AU+Animal based expressions. Lowering the forehead is common between Sadness and Fear which might be the reason behind this confusion.

Table 3.2: AU based expressions confusion matrix

%	Joy	Angry	Sad	Disgust	Surprise	Fear	Neutral	None
Joy	**90.5**	0	0	0	9.5	0	0	0
Anger	0	**80.9**	0	4.8	14.3	0	0	0
Sadness	0	0	**100**	0	0	0	0	0
Disgust	0	0	80.9	**9.5**	0	0	4.8	4.8
Surprise	0	0	14.3	0	**61.9**	23.8	0	0
Fear	4.8	0	14.3	19.0	28.6	**33.3**	0	0

Table 3.3: AU+animal based Expressions confusion matrix

%	Joy	Angry	Sad	Disgust	Surprise	Fear	Neutral	Nn
Joy	**95.2**	0	0	0	0	0	4.8	0
Anger	0	**85.7**	0	0	4.8	9.5	0	0
Sadness	0	0	**95.2**	4.8	0	0	0	0
Disgust	0	0	23.8	**42.8**	4.8	9.5	4.8	14.3
Surprise	0	0	0	0	**71.4**	28.6	0	0
Fear	0	0	47.6	23.8	4.8	**19.0**	0	4.8

3.2 Software Design

The eBear is equipped with several input and output devices: Tablet, Head Camera, Head Servos, Mouth LCD, and the Kinect Sensor. A C# based software was developed in Microsoft Visual Studio environment to control and manage all these devices. Except the screen, all of the other devices are connected to the main PC via a USB connection. A UDP protocol was utilized to connect the tablet to the main PC. When the eBear turns on, it keeps looking for a person to start a conversation. They keep talking as an open conversation until the user asks for other programs such as exercise session. To achieve this goal, several threads were defined as background workers which run in parallel:

- A main thread for the open conversation which is always active.
- A thread for switching to other modes of the interaction such as exercise session which runs when the user asks for it.

- A thread for acquiring the skeleton data from the Kinect sensor which is always running.
- A thread for reading facial expressions from the head camera which is always running.
- A thread for sending/receiving UDP commands to/from the remote screen which is always running.

In the following each software module is explained in details.

3.2.1 Interactivity

One difficult challenge in designing social robots is developing the spoken dialog system and natural language processing. An ideal social robot should be able to completely understand the spoken dialogs while fully interpreting all the visual cues such as the facial expressions or body gestures. There is a huge variety in natural languages which makes the Natural Language Processing (NLP) a challenging task. One of the objects of the eBear was to gather visual and speech information and provide an appropriate response. NLP still is an open research and there is no full language processor available that could be integrated into the robot. Online chatter robots are the closest programs to a full natural language processor (14) (17). Most of the chatter robots or chatbots provide answers using different pattern matching algorithms. Therefore, if their database is huge then the provided answers will be more logical. A rule-based dialog manager was also integrated into the robot. For each of the dialog managers, chatter robot and the rule-based dialog manager, the speech had to be recognized, then processed in the dialog manager and at the end was synthesized using a Text-To-Speech engine. Here each part is discussed separately:

3.2.1.1 Speech Recognition

As it was discussed in the hardware design section, the eBear is equipped with the Microsoft Kinect sensor version 2. The Kinect sensor comes with a microphone array which provides useful features, in particular for robotic purposes such as sound direction detection. We used this microphone array as our main sound input to the system. Microsoft provides a full Software Developer Kit (SDK) for Kinect sensor which comes in different programming languages. The C# SDK was chosen to be compatible with the main software. We used Intel RealSense SDK (8) as the speech to text engine. Intel RealSense SDK is free for non-commercial purposes. Most of the avaiabale speech recognition modules require the user to define a set of words as a dictionary which can be useful in some applications. One advantage of Intel RealSense speech recognition module is that it already comes with a full dictionary and it can be used as a dictation engine. The other good feature of RealSense speech recognition is a set of flags which take care of various speech conditions such as ”Speech started” or ”Volume is high”. We used these flags to notify the eBear when the user starts talking and when the speech is finished.

3.2.1.2 Emotion Recognition

The eBear utilizes both facial and speech emotions of the subject while interacting. These emotions are used to generate more suited facial expressions. In another word, the goal is to make the user feels that the eBear understands his/her emotions. In the following, the facial expression recognition and the speech emotion recognition modules are explained.

Facial Emotion Recognition Intel RealSense SDK provides an accurate facial expression recognition software. It is capable of recognizing the 6 main facial expressions including: Neutral, Happy, Sadness, Surprise, Angry, disgust. By default, the eBear recognizes the facial expression of the subject and replicates the emotion.

Speech Emotion Recognition Aylien (1) is an on-line natural language processing platform. In order to extract the emotion of the speech, their text sentiment analysis API is used. This API takes the speech text as input and provides the sentiment of the speech which can be either of the following: Positive, Negative, and Neutral. It also outputs a confidence level between 0 and 1 showing the accuracy of the analyzed sentiment. When the sentiment of the speech is negative the eBear should feel sad and when it is positive the eBear should feel happy.

In order to combine the results of the facial and speech emotion recognitions, the confidence measure of the speech emotion is utilized. If the confidence level of the speech emotion is less than 0.7, the facial emotion results are used and the eBear replicates them. If the confidence level goes above 0.7, the sentiment is used to make the robot happy, sad, or neutral. The sentiment confidence threshold is adjusted experimentally.

3.2.1.3 Dialog Management

We developed and integrated two different natural language processors for different interaction scenarios into the robot. For the open dialog conversation, we used one of the chatter robots of Pandorabots (17). The chatter robot was not used in any of the private conversations such as the mood evaluation part which will be discussed in section 3.2.3. For the other parts of the conversation, we designed a rule-based dialog manager which was further developed through the interaction sessions. The two dialog managers run in parallel. When the speech recognition module provides a text, in the open conversation part, the input from the speech recognition module is fed into the chatter robot interface and the output is sent to the Text-To-Speech engine. If the input text includes a set of keywords in a specific order which were defined in advance, the related session is started.

3.2.1.4 Speech Synthesizer

Basically speech synthesizer is a Text-To-Speech or TTS engine. We decided to use a male voice with a tonality matching the eBear's appearance. Most of the TTS engines have a robotic like voice which is not desirable for the eBear. Neospeech (6) is a TTS engine which has a wide verity of voices with different genders and accents. We contacted them and they agreed to provide us with a 6-month full academic license to use their TTS engine in the eBear. We integrated their TTS in the eBear but there was a compatibility issue between the rate that speech was prodcued by the TTS and the aligner software. The aligner software is used to align the lip movements with the speech. This issue would happen several times in each session and we had to change the TTS engine. Microsoft provides a few voices which could be used in our software for free. We chose Microsoft Mike which is a mid-age male voice as the TTS engine.

3.2.2 Entertainment

One of the purposes of the eBear was to entertain the participants with various activities to improve their mood and cognitive levels. In (65) they developed a set of computer games as a treatment in geriatric depression (DP). They concluded that computer games can be more effective than depression drugs. In their study, 4 weeks of treatment using the games had the same effect as 12 weeks of using depression drugs. A set of entertaining programs were developed and integrated into the eBear which are discussed in the following.

One of the games was designed based on the ”Catch the Ball” game which was introduced in (65). They provided the general scenario of the game but no image or codes were provided. We used the same scenario but the game was designed in Unity Software from scratch. Figure 3.13 illustrates a screen-shot of the designed game. This game consists of 29 levels. Basically, as it is shown in Figure 3.13, there is a ball moving around the screen. At some points, the color of the ball changes to a target color which the eBear names at the beginning of the game. Once the color changes to the target color, the user is supposed to say ”changed”. If the user say changed before the actual change or 4 seconds after it really changed, the eBear tells them to repeat that level. As the level go higher, the speed of the ball increases and more balls are added to the screen.

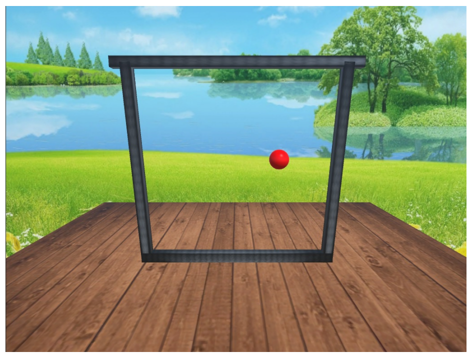
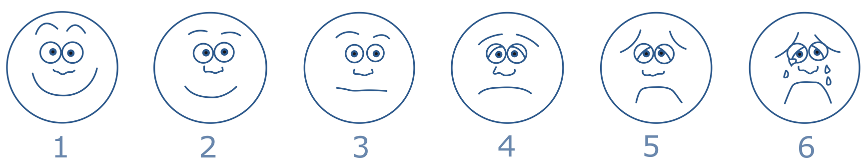


Figure 3.13: Catch the Ball - A screen-shot of the Catch the Ball environment.

Tic-Tac-Toe was another game which was designed and integrated into the robot. Unlike ”Catch the Ball”, the Tic-Tac-Toe game required the user to touch the screen. Tic-Tac-Toe was designed with three levels of difficulty. In each level, the two players play for 5 rounds and whoever wins more times is the winner.

3.2.3 The Face Scale Mood Evaluation

In this section, we talk about a mood evaluation technique which was implemented to be used by the eBear. As it will be discussed in the evaluation chapter, we evaluated the study using different measures. However, one of the goals of the eBear was to be able to automatically assess the mood of the elderly user. Evaluation of the patients condition is a vital part of each psychological experiment. The Face Scale is a nonverbal mood evaluation technique introduced in (61). The advantage of the Face Scale is that this technique is all pictorial and does not need reading literacy. To best of our knowledge, this is the first platform which evaluates the patient mood in this manner. Basically, in this method a set of face pictures with different moods are shown to the participants and they are asked to choose the one which better describes their current mood. In (61), they used 20 pictures of genderless faces. Since the eBear is supposed to run this evaluation at least one time per session, showing 20 pictures in each evaluation would be tiresome. (92) used the same evaluation methodology but with only 7 pictures. The differences between these 7 pictures are more than the ones between the original 20 pictures so that they can cover the whole mood range. We developed a C# based software to perform the evaluation. Each time the robot wants to perform the evaluation, the user is asked to if he/she is willing to tell the eBear about his/her mood. The main program activates the evaluation software which is installed on the screen. The eBear starts with a brief explanation about how the evaluation works. Then the eBear starts showing each of the 6 face pictures on the screen. The pictures are depicted in Figure 3.14. As it can be seen, there is a number below each of the face pictures and the eBear calls the number while showing each of the pictures. At the end, the eBear shows all the pictures in one frame and asks the user which one of the faces better describes your feelings in the current moment. The participant is supposed to say the number of the face. Depending on the selected mood, the eBear provides different responses. For instance, in case of number 1 the eBear will say *"That is great. I am glad to hear that."* and in case of number 6 the eBear says *"I am really sorry to hear that but we should definitely talk more and I promise you will feel better by having me around you."*. The result of each mood evaluation is logged along with the time stamp for further analysis.



(a) **(b)** **(c)** **(d)** **(e)** **(f)**

Figure 3.14: The 6 face pictures used in the Face Scale mood evaluation (11).

4

Get Up and Go Test

Human gait parameters are known to contain significant information about individuals' physical mobility and particularly mobility of the elderly who are at risk of falling. Clinicians have developed methods such as the Get-Up-and-Go Test (GUGT) (69) for gait analysis and screening of the elderly. Such methods are able to predict subjects mobility and determine whether they can walk safely without the risk of falling. Since these assessments are done frequently, there is a great need for developing automatic and inexpensive computer systems capable of performing such assessments in the home of the elderly.

Different technical approaches and sensors have been proposed for gait assessment. One of the common approaches is the use of wearable sensors such as accelerometers or gyroscopes (27), (40). These devices mostly provide accurate gait parameters. Besides, not being expensive, having light weight and being small are other advantages of these sensors. However, these approaches would need a supervisor to help subjects wear the sensors and maintain them frequently. As a result, they are mostly suitable for laboratory purposes and not preferable for the elderly to use them frequently at home (30).

On the other hand vision-based gait analysis systems have received great attention in recent years (70). Some of these systems use regular RGB cameras while some others use more sophisticated sensors such as the Microsoft Kinect. In the camera-based systems, usually a calibrated array of cameras is utilized to provide a 3D representation of the scene. It has been shown that these methods are capable of providing an accurate model of the subject. For example in (58), two calibrated cameras were used to generate a 3D representation of the scene. Using the 3D model, various gait features were extracted including: torso angle, thigh angles and shank angles. However, the need of using more than one camera and their calibration and alignment issues have made it difficult to be used as an in-home screening system.

The Kinect sensor provides RGB images, depth range information and the human skeleton. Using the Kinect sensor for physical mobility assessment was first presented in (79). In that proposed method both a stereo-vision system and a Kinect sensor were used for comparison. In one experiment gait parameters were extracted using the Kinect depth output. In the second part, the same parameters were extracted using the output of the stereo vision system. Finally the results of both systems were compared with the results of a Vicon motion capture system. The experimental results showed that the Kinect sensor measures the gait parameters with a sufficient accuracy.

Another example is the work reported in (38) where the skeleton data of a Kinect sensor was utilized for gait analysis. First some features were extracted from the skeleton data and were fed to a regression model. After that a state machine was used to produce desired states such as whether the foot touches the ground or not. In addition, other features such as the arm kinematics were measured, which show that a wide range of parameters can be extracted from the skeleton data. Nevertheless, they only extracted some features and no classification results were reported by the authors.

Using Kinect sensor eases the gait analysis and extraction of standard stride information with high accuracy in the home environment. Nonetheless, many researchers have focused on only extracting gait parameters for further analysis and there are limited works for automatic classification of subjects' degree of gait severity (40). Hence, designing an inexpensive system that extracts discriminative feature for accurate classification can be useful in alerting patients and clinicians without having a supervisor at home. This chapter presents a methodology for classification of people into two categories, high fall-risk versus low fall-risk, based on their performance in the Get-Up-and-Go Test using Kinect sensor. In our approach, we first use image processing and computer vision algorithms to extract some desired features from the human skeleton data provided by a Kinect sensor. The features include number of steps, average step duration, and turn duration for gait parameters and distance between the elbows, angle between the legs, and angles between the shank and the thigh in each leg (knee angles) for anatomical parameters. Then using a Support Vector Machine (SVM) classifier, subjects are classified into those categories.

4.1 Get-Up-and-Go Test

In terms of physical mobility a person is considered as independent if certain basic skills can be performed without any help of others (46). Get-Up-and-Go Test (GUGT) is a well-known simple test for mobility assessment which consists of basic everyday movements (62). In GUGT a subject sits on an arm chair, gets up, walks a three meter path, turns, walks back to the arm chair, and sits back down. In this test, subjects are asked to perform the task without any help from other people or objects (unless it is necessary), and physical mobility of the subject is rated on a scale of one to five according to the observation of a clinician. The problem with this method is the imprecision of the scoring system. A modified version of this test is called Timed Up and Go test (69). This test computes a score based on the time taken by an individual to stand up from an arm chair, walk a distance of three meters, turn, walk back to the chair and sit down. Due to the timing, this version is more precise than the GUGT in scoring physical mobility. In (67), it is shown that the time score and its variability between test trials correlates well with the physical mobility.

Using the results of the GUGT and the Timed Up and Go test, clinicians can give an estimate of the overall muscle strength and balance of the body which can be used in fall prediction. Generally these tests show whether people are safe on their own or not when it comes to their mobility. At first, each subject sits on an arm chair and is ready to get up. One mobility indicator is the smoothness of getting up. When the subject uses any kind of help to get up or when getting up is not smooth, then it is a sign of abnormality. The next step is to start walking. Any gap between the time that the subject gets up and the time that he/she starts to walk is a sign of stabilizing and is abnormal. When the subject is walking, several abnormalities can be observed such as slow gait speed, feet dragging on the ground, deviating from the straight path, and having severe side to side movements. One significant part of the task is the turning part. Usually, one or two steps are enough to make a complete turn. Using more steps in turning, which corresponds to more turning time, is a sign of abnormality. In addition, some measurements on the anatomical configuration such as the angle between legs, the angles of knees and the distance between elbows, can be indicators of physical mobility while performing the test. In this paper, we extracted some of these abnormality indicators from a Kinect device automatically.

4.2 Automatic Processing of GUGT

In order to measure the human's physical mobility in the GUGT, we used a Kinect sensor to capture a video and track the person's skeleton model while performing the test. The Microsoft Kinect sensor contains of an RGB camera, a depth sensor and a multi-array of microphones. The depth sensor consists of an infrared laser projector and an IR camera with the sensing range of 0.8 meters to 4.0 meters which captures depth images in resolution of 640×480 pixels at 30 frames per second (5). It is capable of tracking the skeleton of one or two people moving within a practical range of 1.2 to 3.5 meters. The provided skeleton consists of 20 joints in the body with 30 frames per second. Fig. 4.1 shows a sample layout and joint indices of the virtual skeleton of the Kinect.

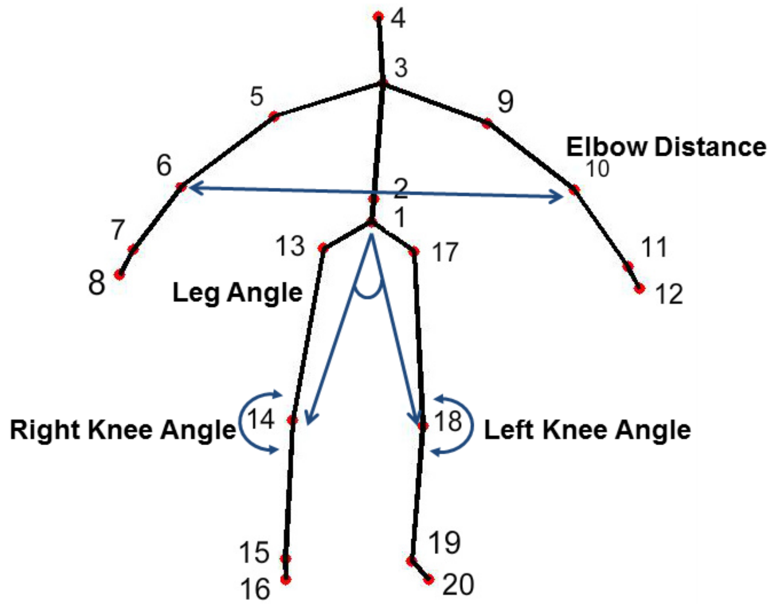
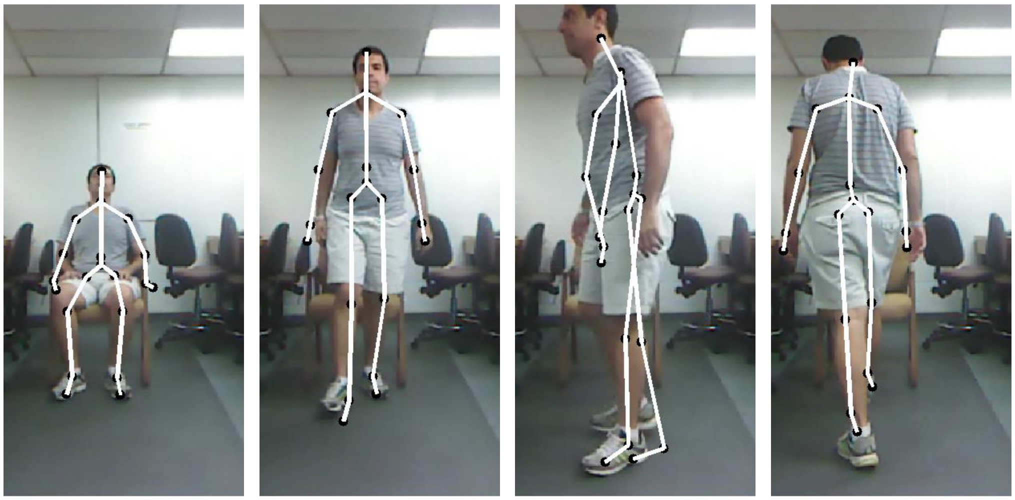


Figure 4.1: Angles - Layout of the Kinect skeleton data

We collected a video dataset of the elderly people performing the GUGT while a Kinect sensor captured a video and tracked the person's skeleton model. Twelve subjects with ages ranging between 65 and 90 enrolled in the study. A geriatric physician reviewed the videos offline and categorized these subjects based on their gait movement and severity of their physical mobility into two categories. The first category includes patients that are relatively safe on their own with low fall risk, and the second one includes patients with high risk of falling that have severe physical mobility issues. The videos were then processed using image processing and computer vision algorithms aimed at extracting some features that were used for classification.



(a) Sitting **(b)** Walking **(c)** Turning **(d)** Walking-back

Figure 4.2: A subject performing the Get-Up-and-Go Test

4.2.1 Feature Extraction

The Kinect sensor detects and tracks the positions (x, y and z) of 20 joints of the human body skeleton. Since the joint points measured by Kinect can be noisy, we apply a median filter of size 5 to remove the noises and improve the accuracy of the measurements. Afterwards, two types of features are extracted: *gait parameters* and *anatomical parameters*.

4.2.1.1 Gait Parameters

In the GUGT a person is instructed to get up from an arm chair, walk, turn, walk back to the starting point, and sit back down. The path is automatically segmented into three phases including: *Seated phase*, *Walking phase*, and *Turning phase*.

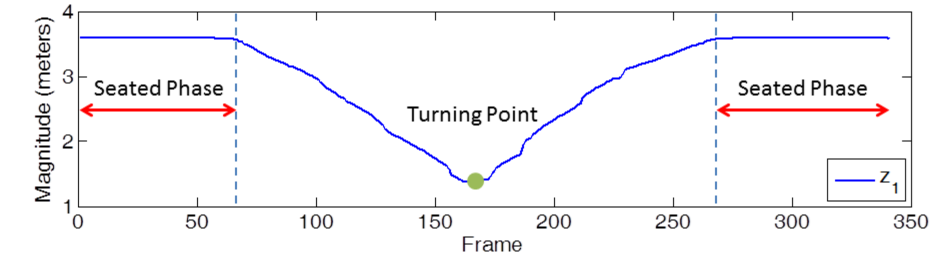
Using the position of the hip joint (point 1 in Fig. 4.1), we can measure the distance of the person to the Kinect in z-direction. When the person is seated, the z-coordinate of the hip joint (z_1) does not change and is at its maximum. By measuring z_1 in consecutive frames from the beginning of the test, we can extract the seated phase. As the person starts walking towards the Kinect, z_1 decreases and when the person is turning, z_1 is at its minimum (we call this point as turn point). Fig. 4.3a shows the position of the hip joint in z-coordinate (z_1), the turn point, and the extracted seated phases.

Since Kinect is mainly designed to track the human body when facing to the camera, it cannot recognize the skeleton of the body well while the person is turning (see Fig. 4.2c). The time that a person is turning can be defined as the time that the person starts to rotate his/her upper part of body to the time that rotation is done and the subject is ready to walk back. We used the absolute difference between x-coordinates of two elbows ($|x_6 - x_{10}|$) to determine the starting and ending frames of a turn. In other word, when the person starts turning this difference decreases and when turning is finished, the absolute difference becomes the same as the value before turn. Measuring this distance in consequent frames before and after of the turn point, we can extract turning phase. Fig. 4.3b shows $|x_6 - x_{10}|$ of a subject, turn point, and extracted turning phase. Extracting seated and turning phases, we can consider the frames in between as walking phases (Fig. 4.3c).

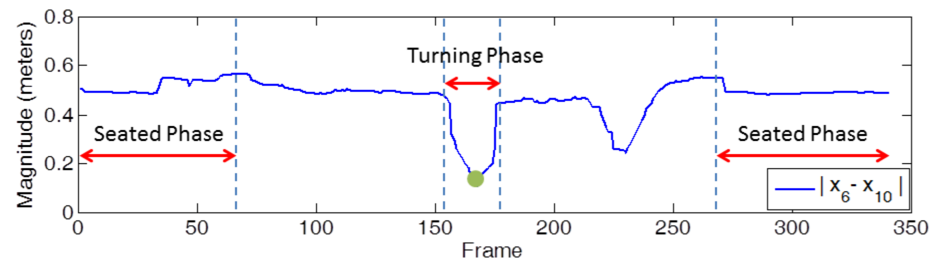
Based on these three phases, three gait parameters are extracted. One of the gait parameters in walking phase is the number of steps that a person takes to perform GUGT. To detect steps, we used the difference between z-coordinates of two heels ($z_{15} - z_{19}$). Extremums of this difference indicate feet being far from each other and zeros correspond to feet being next to each other. Fig. 4.3c shows the difference between z-coordinates of two heels and the extracted number of steps (where the difference is zero).

Duration of each step is another feature which is important in physical mobility measurement. The difference between z-coordinates of two heels ($z_{15} - z_{19}$) gives us the starting frame and the ending frame of each step. For each skeleton data frame, Kinect provides a time-stamp which is used in calculating the duration of each step.

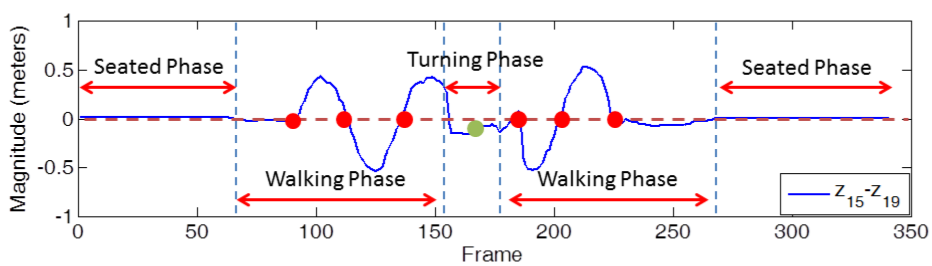
Number of steps in turning phase is another gait related property that contains significant information of the physical mobility. As it was explained before, Kinect cannot



(a) The position of the hip joint in z-coordinate (z_1)



(b) The absolute difference between x-coordinates of two elbows ($|x_6 - x_{10}|$)



(c) The difference between z-coordinates of two heels ($z_{15} - z_{19}$)

Figure 4.3: The extracted seated, walking and turning phases in a GUGT

recognize the skeleton of the body while the person is turning. One approach to overcome this issue is to use another Kinect to look at the subject from side. This approach demands synchronization between Kinects which is against the simplicity of the proposed method. The other approach is to measure the turning time instead of counting the number of steps. We extracted the turning phase automatically and the turn duration is calculated based on the time-stamps of the starting and the ending frames of the turning phase.

4.2.1.2 Anatomical Parameters

The second type of parameters for physical mobility assessment is related to the anatomical configuration of the person while performing GUGT. Various anatomical measurements can be obtained to assess the condition of the body. One feature is the distance between two elbows (joints 6 and 10). The angle between the legs is another feature that is used here. This angle is defined as the angle between two vectors connecting joint 1 to 14 and 18. The other extracted features are the right and left knee angles that are the angles between the shank and the thigh in each leg. This angle is defined as the angle between two vectors connecting joint 18 to 17 and 19 for the left leg and the angle between two vectors connecting joint 14 to 13 and 15 for the right leg. These features are shown in Fig. 4.1.

4.2.2 Classification

The features described in the previous section are used in classification of gait via a soft margin C-SVM classifier. The gait parameters provide us three numerical features for each sample which are: (1) The number of steps that have been taken to perform the test; (2) The average duration of steps in seconds; (3) The turn duration in seconds. These numbers can be fed into the classifier directly.

The anatomical parameters can be easily measured in each frame of the skeleton data, but to be used as the input of the classifier, they should be comparable. In other words, there is no guarantee that subjects finish the task in the same amount of the time and hence there are different numbers of features for each subject in each test. To overcome this problem we have used the Bag-Of-Words (BOW) (34). The Bag-Of-Words or Bag-Of-Features is a simple approach that is commonly used in visual object classification, text categorization, etc., where there are different number of features for different samples. Assume there are $\{N_1, N_2, ..., N_k\}$ various number of features for k samples, this model represents those features in M keywords where in this paper $M \ll \{N_1, N_2, ..., N_k\}$. For this purpose, first a clustering algorithm is used to cluster all training samples features to M clusters. Each cluster is known by its center. Each feature of each sample is assigned to one of these centers, and then for each sample, the histogram of the features in these clusters represents new M dimensional features of the sample. This M dimensional output along with the three gait parameters are fed into the classifier.

4.3 Experimental Results

The Kinect sensor was mounted on a table with an approximate height of 120cm. As the maximum practical range of Kinect sensor is 3.5 meters (5), and we want Kinect to view the whole body during the test, the arm chair was located at a distance of approximately 3.5 meters to Kinect and subjects walked a path of approximately two meters instead of three meters in original GUGT. Subjects, with their normal clothes, were instructed to first sit on the chair, stand up and start walking for two meters, then turn in place and walk back to the chair and sit back down. They were asked to walk completely normally during the test.

We measured the accuracy of the proposed classification technique on the dataset of 12 elderly patients collected at the clinic. Among those 12 patients, five subjects are categorized as low risk of falling and seven of them had high fall risk by an expert physician. Each patient repeated the test three to six times (each time is called a sample). A total of 50 samples were captured. We have used K-means clustering to cluster the anatomical parameters for the Bag-Of-Words and a C-SVM with radial basis kernel for classifying. K-means requires the number of clusters (K) to be defined which corresponds to the number of words in BOW. We evaluated the algorithm using 4 to 24 clusters, and it turned out that 10 clusters gave us the best classification rate.

To measure the performance of the proposed classification, we used a leave-one-subject-out technique. Particularly, the classifier is trained with samples of all subjects except one subject. For testing, the classifier classifies all samples of the left out subject. Every time one subject is left out and the training and testing procedure are repeated till all subjects are covered. Finally, the accuracy of the classification is reported as the average of samples being classified correctly. As K-means takes random initial clusters, BOW may generate different features each time which may affect the classification performance. We repeated the whole procedure ten times and on average the classification accuracy is 67.40% with standard deviation of 4.72%. Table 4.1 shows the average confusion matrix of the classification.

Table 4.1: Classification Confusion Matrix

	Low Risk	High Risk
Low Risk	**16.2**	7.8
High Risk	8.5	**17.5**

Results suggest that the classification method and features we designed provide an effective means of distinguishing patients who have a high risk for falling from patients with a lower fall risk. This inexpensive, easy to use, Kinect sensor-based approach can easily be used in the subject's home or by lower skilled healthcare personnel and relayed to a physician to further investigate as appropriate.

In this study, we extracted only three gait parameters. Using depth data as well as skeleton tracking, we can extract more complicated gait parameters indicating abnormality in physical mobility such as smoothness of getting up, using any kind of help to get up, and feet dragging on the ground.

5

Activity Assessment

In this chapter we present two methods for coaching users through exercise. As said, it has been proven that lack of mobility is a major sign of depression and this project aims at developing methods for minimizing this factor as well. We want the companionbot platform to engage users in conversation and prompt them to exercise with the aim of increasing their mobility. The user should be monitored while performing the exercise and the related feedbacks should be provided to keep the user engaged and motivated. The Microsoft Kinect sensor is used as the eyes of the system to constantly monitor the users activities. In case where the user does not move for a period of time, the robot starts a conversation with the user. During the conversation, the user is prompted to start some pre-defined exercises. The proposed algorithm analyzes the users movements and provides positive feedback to correct the movements as well as keeping the user engaged. The feedback is provided during the exercise. Toward that goal, this chapter addresses the computer vision and feedback algorithm needed for the robot.

As mentioned before, we aim to take the companionbot to the house of elderly people so that it becomes a companion, be trusted and treated as a member of the family and eventually improves their quality of life. The robot will be placed in a corner of the main hall of the house, where the user spends most of the day. It has been assumed that in such a position, the Kinect sensor will see the user most of the time. In the designed scenario, when the person is in the view angle of the Kinect sensor, he/she is tracked all the time. When the Kinect detects that the user has not moved for a period of time, the robot starts a conversation with the aim of motivating the person to move. One way of prompting the user to move is to ask them to perform an exercise. If he/she accepts to do the exercise, the system will monitor the movements. Using a vision algorithm, movements are analyzed and according to the quality of their moves, appropriate feedback is provided. Utilizing feedbacks not only corrects the way users perform the exercise, but also keeps a conversation between the robot and the user so that the user remains engaged. Because this system is mainly designed for the elderly people, not all types of exercise can be asked to be performed. Chair exercise (23) is a type of movement which is mostly recommended for elderly people and is used in our system. This type of exercise decreases the risk of falling caused by balance problems. Also, it has been proven that such exercise has many positive effects on their bodies including lubricating joints, strengthening muscles and increasing the blood circulation (29). Method 1 that is presented in Section 5.1 was designed and developed to assess the quality of exercises. However, since this method uses low level features for evaluation, it requires the user to follow the exercise pattern and if they perform an additional movement, the system might provide inaccurate assessments. The second method (presented in Section 5.2) is our modification of the algorithm proposed in (68) which utilizes high level features and is more robust against input variability and noises.

5.1 Method 1

5.1.1 Vision algorithm

The vision algorithm of the system consists of several steps. To create our model for a specific exercise, we first record a trainer while performing the desired exercise in front of the Kinect several times. Then the input stream is preprocessed which consists of angle calculation and filtering. After that, the input is segmented using a Support Vector Machine; each segment includes a single exercise, which has been repeated several times. Finally, Dynamic Time Warping (74) is used to compare each segment to the reference model.

We use a Kinect V1 sensor placed on the body of the robot to monitor and capture the physical movements of the user. The version one of Microsoft Kinect sensor consists of a depth sensor, an RGB camera and a microphone array. The depth sensor uses an infrared laser projector and an IR camera to capture depth images of 640×480 resolution at 30 frames per second. The sensing range of the Kinect sensor is between 0.8 meters to 4.0 meters (5). Because of this limitation, we assume that the robot is within this distance range to the user. In addition to the depth image, the Kinect can detect the skeleton of the person in front of it. The skeleton output contains the positions (x, y and z) of 20 joints in the body with respect to the Kinect coordinate system. Fig. 5.1 depicts the layout and joint indices of the skeleton of the Kinect.

5.1.1.1 Preprocessing

Since the joints' positions measured by the Kinect can be noisy, a median filter of size 5 is applied to the input data to remove noise and improve the accuracy of the algorithm.

To analyze the skeleton, three possible measures can be utilized: absolute position of the joints, relative position of the joints, or the angle at each joint. On the other hand, we

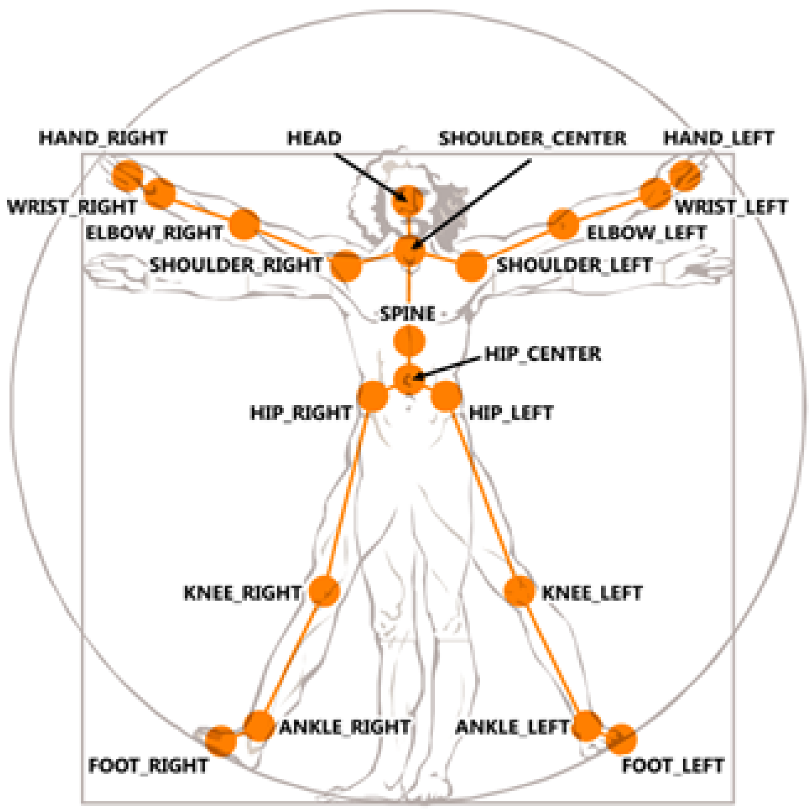


Figure 5.1: Kinect Layout - Layout of the Kinect skeleton data (5)

need the measure to be independent of the body shape and measurements of the user. In addition, in the Kinesiology's standard, the position of each bone is determined using its angles in relation to the reference planes which are sagital, frontal and horizontal (88). For these reasons, we use the spherical coordinate system to represent each bone. In this system, each bone is represented using a polar angle (θ) and an azimuthal angle (φ). As explained before, positions of joints are determined with respect to the Kinect coordinate frame. Because we need the angles to be independent of the direction of the person in relation to the Kinect, a new coordinate frame was defined with respect to the skeleton itself. After analyzing different skeleton data, we noticed that joints number 3, 5 and 9 are the ones with the least amount of noise and error. Using these points a new coordinate frame was defined with its origin on the HIP point of the body (joint 3).

Based on Fig. 5.1, it can be seen that the Kinect defines 19 segments in the body. Because we aim at chair exercises, we are only interested in the ones that define bones of the arms and legs. We further simplify and assume that only arms are utilized in the exercises. However, it can be easily extended to all segments of the body. There exist three segments in each arm which result in 12 angles in total.

After the conversion, all the angles are smoothed using a first order Butterworth low-pass filter with a normalized cutoff frequency of 0.9. Fig. 5.2 shows the data of one angle of the reference model of a particular exercise. It also illustrates the same angle data of a user while performing that model exercise with different speeds and intensities.

5.1.1.2 Segmentation

We do not limit the way that the users should perform the specified exercise. They may perform the desired exercise as many times as they wish and there is no need to perform each repetition consecutively. Because of this assumption, the segmentation problem can be divided into two stages. First we need to detect whether the person is doing the desired

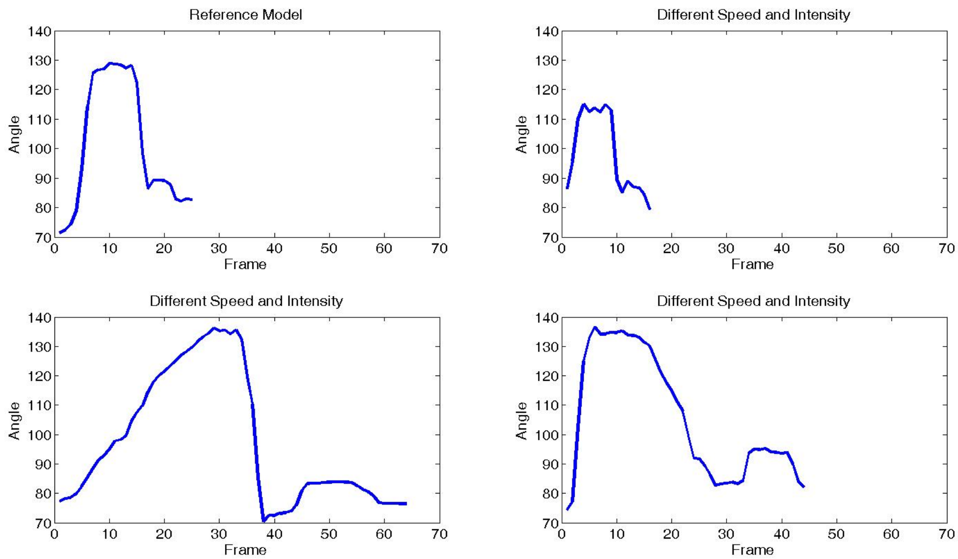


Figure 5.2: - Reference Model and its variations in speed and intensity

exercise or not. After that, in case of repeating an exercise several times, the input stream should be broken down into each individual repetition so that the comparison could be done more accurately.

For the first stage of the segmentation, two assumptions are made (the results show the validity of each). First we assume that when the person is performing, the majority of the angles are in a specified range. In the training phase, we calculate the minimum and the maximum of each angle as $angleMin$ and $angleMax$. Because of the first assumption, we extend each side of the interval by 20% of the range. The second assumption is that for each type of exercise, the ratio of each angle to the rest can be considered as an identifier of that specific exercise. These ratios are calculated for each angle in the training phase and their minimums and maximums are stored. For 12 angles, there are 132 ratios. Each ratio's range is extended by 0.2.

These two identifiers result in 144 thresholds ranges. The thresholds are applied to the input stream. For each frame, if at least 70% of the thresholds confirm the existence of the desired exercise, it will be assumed that in that frame the exercise is being performed. A mode filter of size 10 is used to filter out unexpected noise in the output of the thresholds. Basically, this part defines a mask on the input stream which locates the moments that the exercise is being done.

After applying the mask, we should detect each exercise in the input stream. One exercise can be performed at different speeds or various intensities. For instance, in a simple hand clapping gesture, one may clap faster than he/she is supposed to clap or open his/her hands more than another person. On the other hand, the start index of each exercise is unknown. These variations have made the matching problem challenging. Fig. 5.2 illustrates a few variations of an original template. It can be seen that variations can happen in any order with any intensity. To overcome this problem, a classification technique is used. For classification, we determine parts of the input signal that represent the beginning of an exercise, and parts that correspond to the ending of the signal. For this purpose, each angle signal is windowed into 6-frame windows and each window shares 1 frame with the previous window. In case of using 6 segments or 12 angles of a hand (It varies with the number of segments), we define 450 features for each 6-frame window to make sure that we have utilized various aspects of the input stream:

$Angles$: Because of the fact that angles are person-independent, they are used as a part of the feature vector. This part results in a feature vector of size 6 for each angle.

$Ratios$: For this set of features we calculate the mean of each angle in each window and also calculate the ratio of each one to the rest. The same as the first stage of segmentation, this part produces 132 features.

PCA : This feature is the projection of all 12 angles onto the first principal component of those angles which produces a feature vector of size 6.

$Statistical features$: Difference between mean and the maximum of the segment, minimum, maximum, mean, median, variance, standard deviation, root mean square, first derivative, and peak-to-peak amplitude of the segment. This part results in a feature vector of size 10 for each angle (39).

$Morphological features$: values of morphological points, time of the morphological point, first derivative and second derivative. This part results in a feature vector of size 10 for each angle (39).

In the training phase, first the training data are windowed into 6-frames ones and we labeled each of them as ”Begining”, ”Ending” or ”Middle”. Then for each window, 450 features are computed. Two linear Support Vector Machines (SVM) are used to classify each 6-frame window into either ”Beginning” or ”Ending”. A mode filter of size 5 is used as a simple post-processing technique to decrease any possible false prediction.

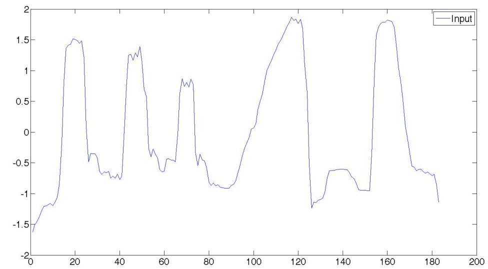
Fig. 5.3b and fig. 5.3c show the result of the classification method applied on fig. 5.3a. It should be noted that only transitions and their directions are of interest. In fig. 5.3b it can be seen that even after applying the post-processing method, some transitions are falsely detected. To make the algorithm robust against these false predictions, an extra step is taken to complete the segmentation part. In this step, we define three events that must happen consecutively, so that a segment can be detected as an exercise. Each exercise starts with one transition from zero to one in the output of the ”Beginning” SVM and ends with a transition from zero to one in the output of the ”Ending” SVM. But the algorithm looks for the ”Ending” transition after a transition from one to zero happens to the ”Beginning” SVM output. This part rejects any other transitions that may happen as a result of error in the predictions. Fig. 5.3d illustrates the segmented areas.

5.1.1.3 Comparison

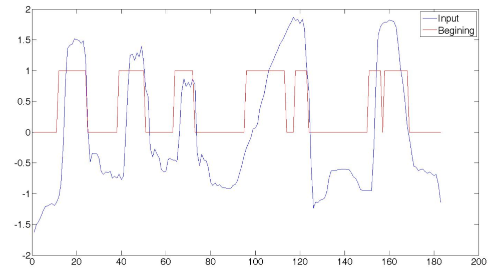
Now that we have all the exercises sub-actionssegmented, they have to be compared to the reference model. It should be mentioned that we aim at elderly people, in particular people with depression. Because of this, we do not need to provide a detailed comparison of the movements. However, our method of comparison provides more details than we need for our purpose.

This comparison should be done with respect to two criteria: speed and intensity. In the first type we evaluate whether the user is exercising fast, slow or at a normal pace. We have the reference model and the number of frames it takes to be performed. The segmentation section gives us the starting and ending points of each exercise. Using these, we can calculate the number of frames of each performed exercise. For comparison, we define an interval around the model's number of frames. If the number of frames of the performed exercise falls within this range, then the exercise is categorized as normal pace. Similarly, if the number of frames is less than or more than the defined range, the exercise will be considered as fast or slow, respectively. The accuracy and sensitivity of the comparison depends on the size of the defined interval.

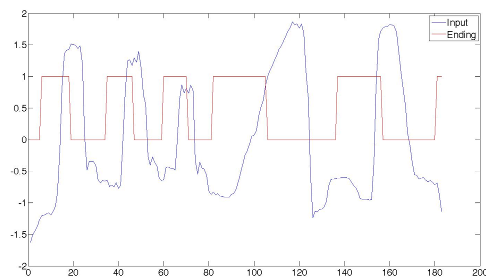
Dynamic Time Warping (DTW) (24) is utilized to compare each performed exercise with the reference model in terms of intensities. In terms of intensities, the mean of each angle is compared with its reference to see whether it is smaller or bigger than its model. DTW finds the optimal path between two time series, which has been used for many purposes including: classification, similarity measurement and finding the related segments of two time series. Fig. 5.4 illustrates the idea behind DTW. Using this technique each point of one of the time series is connected to the most similar point in the other time series. After applying DTW on all the detected exercise segments, the corresponding areas are recognized. According to the required accuracy, different evaluations can be performed. In our case, because of dealing with depressed elderly, we do not want to provide detailed feedback. Due to this, first the differences between the unwarped version of the reference and the exercise segments is calculated. Then we apply two thresholds on the magnitude of the difference and the number of points that have that difference. If the final threshold output is bigger or smaller than a specific value then the algorithm decides that the corresponding angle should be decreased or increased respectively. The specific value is determined through experiments.



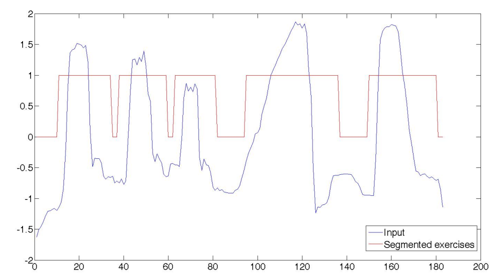
(a) The input



(b) Detected beginning points



(c) Detected ending points



(d) Segmented exercises

Figure 5.3: The output of the classifier

5.1.2 Decision algorithm

It needs to be noted that this decision algorithm is different from the final algorithm which was used in the field trial. This simple decision algorithm was designed toward that ultimate goal so that we can assess people's opinions about the robot and modify the algorithm based on their ideas. Fig. 5.5 depicts the flowchart of the proposed algorithm. A speech recognition module is developed in order to process the feedback of the users. Because we only want the user to agree or disagree with the spoken sentences, the module only detects ”Yes” and ”No”.

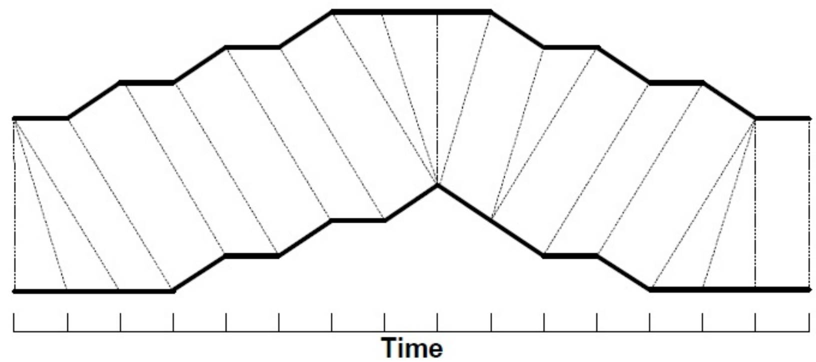


Figure 5.4: Dynamic Time Warping (75)

In SAR, because robots have to make decisions and speak with people, one of the concerns is the repetitiveness of their language. In order to decrease the repetitiveness in the robot's feedbacks, for each comment, various equivalent sentences are selected. Each time, one of the sentences is selected randomly. According to (33), this randomness has a positive effect on the user's perception of the robot's intelligence. In addition, special attention was placed on the type of sentences and the way that they make the person feel. The user should see the robot as a friend and a companion not as an extra tool, which might make them feel nervous or under pressure. For this purpose, all the sentences, even the negative ones have a sense of friendliness. Another point which is taken into account is the facial expression of the robot. Based on the type and meaning of the feedback, the robot shows appropriate expressions.

According to the described scenario and the flowchart, the robot is placed in a corner of the target house. First the robot looks for the person and if no one was within its view angle, it will keep searching until a person is detected. We assume that there exists only one person in the room. Once a person is detected it waits for a number of seconds, which varies from person to person and is determined practically. After that amount of time, the robot starts encouraging the user to exercise. In case the user rejects the exercise offer, the robot will totally agree with that decision and will not force the user at all. The user is assumed to be aware of the type of the exercise in advance. If the user gets motivated to exercise, the robot will coach him/her and after finishing the exercise, appropriate comments will be provided. Then the user is asked to perform the exercise again. This process will continue.

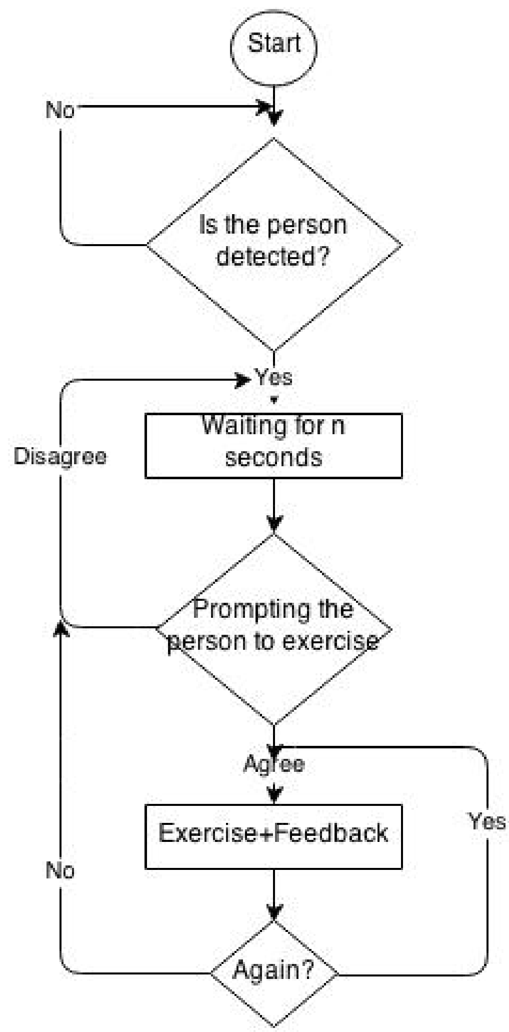


Figure 5.5: The decision algorithm of the proposed system

5.1.3 Evaluation and Discussion

In order to evaluate the proposed system, two experiments are designed and tested: User feelings about the robot, and the accuracy of the vision algorithm.

5.1.3.1 User Feelings Towards the Robot

Although the robot is designed to deal with elderly people, this part experiment was designed to understand general users' feelings toward the robot. The aim is to apply their comments on the proposed system in order to improve that. To do so, 9 participants (3 females and 6 males) aging from 22 to 45 years old (*mean*=30.33) were asked to test the proposed platform. In each session, one participant was in one room with the robot and the whole procedure took around 15 minutes depending on the user's answers to the feedback. They were given a questionnaire to rate the robot from various viewpoints. Below is the list of the questions:

How naturalistic is the robot?

How intelligent is the robot?

How helpful is the robot?

Would you like to have such a robot in your house?

Is it boring?

They were asked to rate each of these questions from 1 to 5. For the first four questions, 1 represents ”Not at all” and 5 means ”Very”. In the last question 1 means ”It is not” and 5 means ”It is”. Before each session the meanings of these questions were explained in detail so that all the participants had the same understanding of the questions. The first question aims to assess whether the robot's appearance and actions are viewed as a cartoonish character or a simple mechanical device, which convey no feelings. Being intelligent means that the robot's decisions and actions are similar to those of a human. The robot is helpful when its interaction with a participant is perceived to make him/her do the exercises and eventually increase their overall mobility. Participants also were asked if they would like to have such a robot in their own houses or not. The last question asks if the robot can be perceived as repetitive or predictable.

Figure 5.6 summarizes the users' responses to survey questions. The results are promising in three of the cases. The users have rated to robot's intelligence with a score of more than 4 which illustrates the way it provides feedbacks is smart. A score of more than 3.5 in helpfulness suggests that having such a robot in a house will get the person to have more physical mobility. An interesting result is that most of the participants would like to have such a robot in their house. A smart robot is always fascinating, hence this case might raise some ambiguity because it is not clear whether their opinion is because of an exciting robot or the fact that it gets them to exercise. This case can be investigated in more detail in future work. The average rating for being naturalistic is approximately 2.5 which shows either its appearance and mechanical features or the way the robot speaks with the user is not natural. The robot was also rated to have a score of approximately 2.5 in boringness, which suggests it might be repetitive or predictable for the user. As discussed, the robot has various versions of each answer and each time one of them is randomly selected, however it seems this in not enough. In future work, other modules, like telling the news or showing some TV shows, will be integrated to the system to add variety.

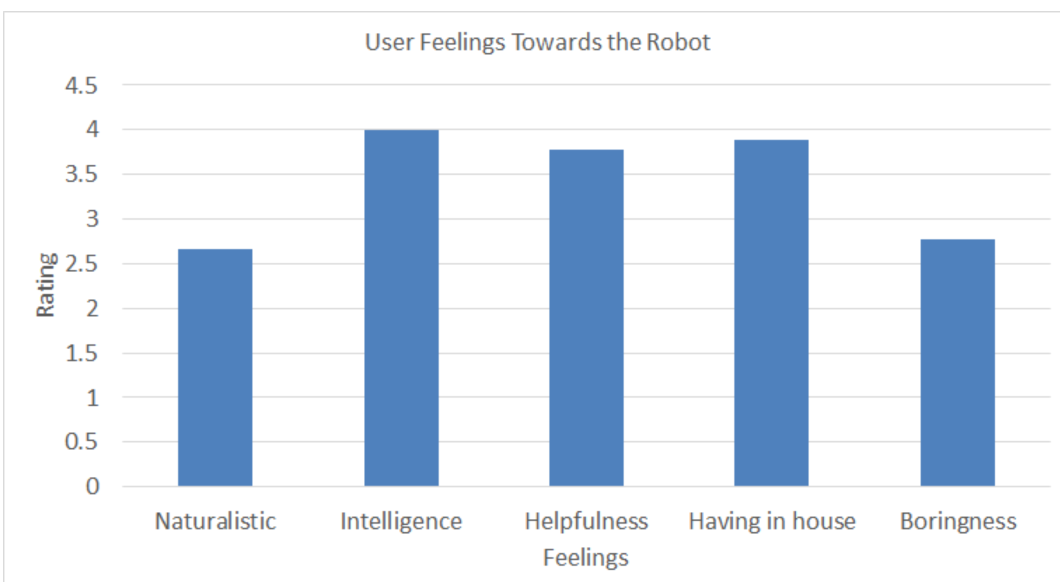


Figure 5.6: Graph of the users' ratings about various aspects of the robot

5.1.3.2 Vision Algorithm

The data of all the 9 participants were recorded and used in the evaluation part. In addition, 2 other users were recorded while performing the same exercise but in different ways including various speeds and intensities. In total, 74 instances of the same exercise segment were captured and used in the evaluation. The evaluation is done in two parts, first the accuracy of the segmenter is analyzed, then the comparison stage is evaluated.

Segmentation Results To form the ground truth, the indices of the starting and ending points of each exercise segment in all the captured videos were manually extracted. In order to calculate the accuracy of the segmentation, three notations are defined the same as ref18: "*Traditional* means the predicted segment should overlap with the ground truth". "*Close* means both boundaries of the predicted segment should fall within 20 frames of the corresponding ground truth's boundaries". "*Tight* means both boundaries of the predicted segment should fall within 10 frames of the corresponding ground truth's boundaries".

Table 5.1 illustrates the accuracy of the segmenter. All the predicted segments overlapped the corresponding ground truth segments, but in 5 cases boundaries of the predicted segments fell outside 20 frames of the corresponding segments. In addition, 14 of the cases did not pass the *Tight* definition.

Table 5.1: Segmentation Results

	Accuracy
Traditional	100%
Close	93.24%
Tight	81.08%

Comparison Results To compare the pace of the detected segments, three metrics are defined: *Normal* means the number of the predicted segments' frames falls within 70 frames of the reference model's number of frames. *Fast* means the number of the predicted segments' frames is less than the reference model's number of frames by a difference of more than 70. *Slow* means the number of the predicted segments' frames is more than the reference model's number of frames by a difference of more than 70. The number 70 was determined practically by the trainer. Because speed of each segment is defined based on the difference between its starting frame and ending frame, the segmentation results also reflect speed comparison results as well. But to be more specific, in 97.29% of cases the speed was determined correctly. Table 5.2 presents the confusion matrix of the speed results.

As discussed in the vision algorithm section, for each segment two thresholds should be calculated. They were determined practically, since it is the trainer's opinion whether a person is performing correctly or not and the threshold can be modified according to the type of exercise or the person. For the intensity comparison, all the videos were reviewed by the trainer and for each segment, a label of "*Good*" or "*Not good*" was assigned, which formed the ground truth. Using the thresholds and the ground truth, in 77.02 % of the cases the algorithm predicted right. Because each exercise includes various parts of a body, it is difficult to come up with perfect criteria for each exercise. The result can be easily changed by modifying the thresholds, which means more flexibility will be given to the way an exercise should be performed. Hence, this accuracy is not of much importance to this system.

Table 5.2: Confusion Matrix of Speed Comparison Results

	Predicted		
	Slow	Normal	Fast
Slow	100%	0	0
Normal	0	95.65%	4.54%
Fast	0	2.70%	97.30%

5.2 Method 2

In the first method all the features were low-level hand-crafted. This made the method not robust against the changes and noises of the input. If the user follows the order if exercise, the algorithm detects all the different segments and provides accurate feedbacks. However, since the robot will deal with elderly people, we cannot expect the elderly user to carefully watch the instructions and perform them in order. (68) introduces a general framework for assessing the quality of actions. They chose two different Olympic sports to show the generality of their method. Since they used game videos for this purpose, we had to modify the algorithm to work with the skeleton output of the Kinect sensor. Here we briefly describe their method and then will explain how the modified version works in details. The input to the algorithm is a video of an action. They extract two sets of features. The first set is low level feature which is extracted using the hierarchical feature described in (56). For the high level features, they ran Flexible Parts Model (95) on the sport videos to calculate the pose of the human performers. They applied DTC on the normalized version of the joint positions and applied a low pass filter to get rid of the noises. They trained a linear Support Vector Regression (SVR) on the result of previous parts to predict the judge scores. For the feedback part, using the mathematical formulation of the SVR, they calculated the gradient of the score with respect to any of the joints. The maximum gradient corresponds to the joint and the direction where the performer should move to get a better result. Since Flexible Parts Model provides the joint positions in 2D space, we had to modify the algorithm so that it could be used with the skeleton output of the Kinect sensor. In addition, other modifications were also made to increase make the algorithm more customizable which will be explained in details.

5.2.1 Vision Algorithm

5.2.1.1 Preprocessing

The authors of(68) applied Flexible Parts Model (95) to extract the joint positions. We use the Kinect sensor for this purpose. The skeleton output of the Kinect version 2 is depicted in figure 5.7. As it can be seen, it can track up to 26 joints whereas the version 1 of Kinect tracks up to 20 joints. Depending on the type of exercise, different joints might be involved in the exercise. This algorithm describes a general framework which can be applied on any set of the joints. In the algorithm description, for the sake of generality, we assume all the joints are involved. We assume the performer is in the sight of the Kinect sensor and all his/her joint positions are being tracked. In case some of the joints are not in the sight of the sensor, Kinect provides an estimation of the positions. For each joint we have the corresponding x, y, and z. Kinect version 2 provides more accurate tracking than the version 1, however still there are some noises in the input data which needs to be filtered. We used a median filter of size 3 which means for every x, y, or z the value of the previous three frames as well as the next three frames are taken into account. Since there is much variety in people's heights as well as the size of different limbs among different people, the position of joints are normalized relative to the head positions. Joint number 21 is the head in Kinect version 2:

$$x\prime^{(j)}(f) = x^{(j)}(f) - x^{(21)}(f)$$

$$y\prime^{(j)}(f) = y^{(j)}(f) - y^{(21)}(f)$$

$$z\prime^{(j)}(f) = z^{(j)}(f) - z^{(21)}(f)$$

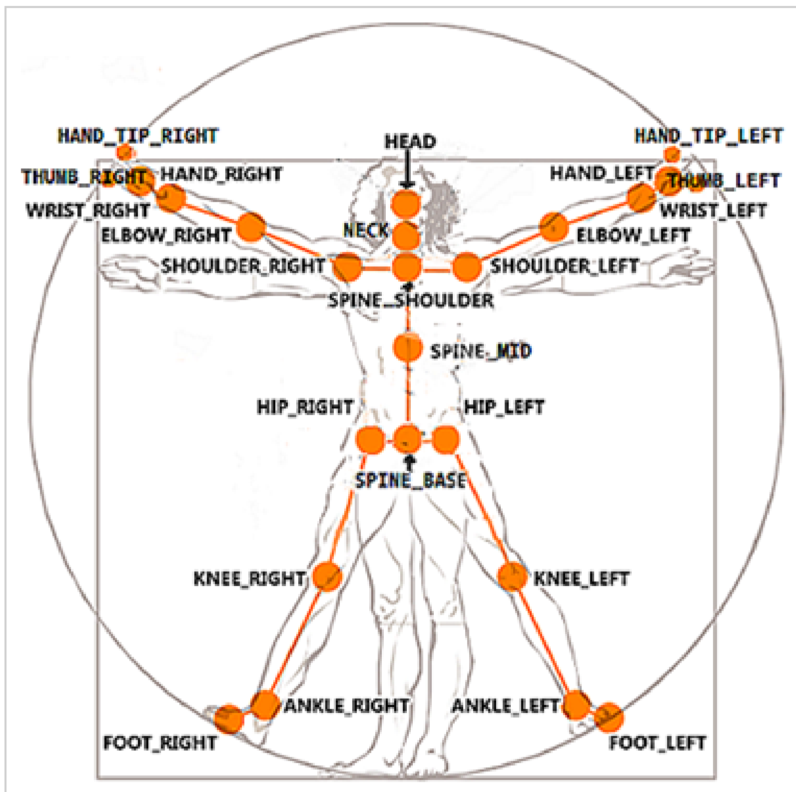


Figure 5.7: Kinect V2 Layout - The skeleton layout of Kinect v2 (5)

Where f is the frame number and j is the joint number. At this point, the joints of interest can be chosen and the unrelated ones can be discarded.

5.2.1.2 Feature Extraction

The joint positions are functions of time and we want to represent them in frequency domain. To do so, we apply Discrete Cosine Transform (DCT) on the joint positions to transfer to frequency domain. Matrix q is created by putting x, y, and z of all joints as its rows. DCT is applied as follows:

$$Q = Aq$$

Where A is the DCT transformation matrix. In order to further remove the noises, the first k rows of the matrix A is chosen. This basically acts as a low-pass filter that removes the high frequency components which are most probably related to the input noise. After applying DCT and the low-pass filter, the rows of Q are concatenated to form the final feature vector. (68) argues that even though DCT and Discrete Fourier Transform (DFT) are close to each other, but DCT performs a better job in this case. They believe that this is because of the fact that DCT provides a more compact representation of the signal. In addition, DCT provides real number which as opposed to the complex numbers of DFT results, less information is lost during the absolute value operations.

5.2.1.3 Learning

The final feature vector is of size $k \times n$ where k is the number of low frequency components and n is the number of selected joints. For each video there will be a score y indicating the ground truth score. The training phase will be done in a supervised manner using Linear Support Vector Regression (L-SVR).

5.2.1.4 Feedback

Providing feedbacks on how to improve the quality of the activity is the other part which was proposed in (68). Basically, L-SVR calculates a weight matrix which will be applied on the input feature vector to produce the output score. The dot-product equation is as follows:

$$S = \sum_{f=1}^{k} \sum_{j=1}^{n} W_{fj} \phi_{fj}$$

Since there is a mathematical relation between the input feature vector and the output score, we can calculate the gradient of the score with respect to each of the input features. In this way, by finding the maximums and minimums of the gradient, we can calculate the joints which were responsible for increasing or decreasing the score the most. In (68) they calculate the gradient as follows:

$$\frac{\partial S}{\partial p^{j}(t)} = \sum_{f=1}^{k} A_{ft} W_{fj} \cdot sign(\sum_{t\prime=1}^{T} (A_{ft\prime}(p^{(j)}(t\prime) - p^{(21)}(t\prime))))$$

Then they find the maximum of this equation and which provides the joint number as well as the direction which that joint must move in order to increase the score. In our algorithm, we used the same gradient approach. However, instead of calculating the maximum of the gradient, we experimentally calculated a set of thresholds for all the joints. For each joint, if the corresponding gradient is more than the related threshold, then it means this joint should move in the direction that the gradient shows. Using this method, we are able to provide feedbacks on more than one joint. To summarize, here are the modifications which were applied on the algorithm that was described in (68):

- The algorithm is designed on the 3D position of the joints.
- Depending on the exercise, only a few of the joints were selected for the calculations.

- A filter was applied on the input raw data.
- The feedback method is changed so that feedbacks can be provided for all the joints.
- We apply the algorithm on real-time data as opposed to a stored database

The next section explains the implementation details as well as the data collection part.

5.2.2 Implementation

(68) has released the source code of their algorithm which is written in Matlab Software. In their paper, they talk about a set of low-level features, but low-level feature extraction was not included in the source code. The code includes the feature extraction and the training section, but not the feedback part. We first modified their Matlab code to meet our requirements. All the trainings and testings were done in Matlab. Then the code was rewritten in C# language in the Microsoft Visual Studio environment to be integrated into the eBear.

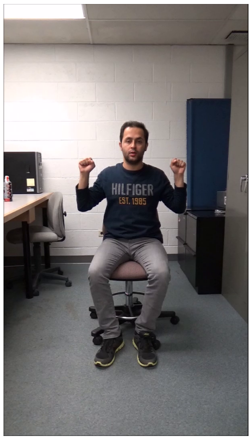
A C# software was developed to record the skeleton output of the Kinect sensor. We setup a laboratory setting in the Computer Vision Laboratory at University of Denver. Three exercises were selected so that it adds variety to the program. The exercises were all chair exercises to omit the risk of falling and only shoulders and arms were involved in all activities. It was assumed that the scores are between 0 and 100. Since a regression model is used to give the scores, the training data have to include wide range of scores to better model the activity. To achieve this variety, for each exercise, three people performed the exact activity for 20 times. The score of these exercises were the maximum which was 100. Then they were asked to perform random movements in front of the sensor. Different segments with various lengths were chosen from the random movements to increase the number of samples. Three L-SVR were trained using these samples. The calculated parameters of L-SVR were stored as an input to the C# program for activity assessment.

One of the problems of the method 1 5.1 was the segmentation. Since DCT is applied on the input joint positions, even if the segmentation is not exact, the algorithm still would provide a score and the related feedbacks. However, to increase the accuracy of the system, the eBear was programmed to guide the user to perform one repetition at a time.

5.2.3 Exercise Session

In this section the scenario in which the exercise is instructed is explained in details. Note that for each sentence, there are a few replacement which are chosen randomly. This randomness decrease the probability of boringness of the program. Hereafter, for each instruction, we only show one of the sentences. Three different exercises are programmed so that the user chooses from them. Figure 5.8 depicts the major parts of these exercises. The exercise section is activated in three cases: if the user asks for exercise, if the operator push the exercise button, or if the user does not move for a period of time the robot activates the exercise. Except the case when the user asks fot the exercise, in the two other cases, the robot starts with asking if the user is willing to perform some exercises: "*Do you want to have some exercises?*". If the user rejects the robot will not insist on it: "*Sure. That is totally fine. Please let me know if you wanted to exercise later.* ". The robot starts the exercise session by asking the user to choose among the three different exercises that are being shown on the screen as illustrated in 5.8.

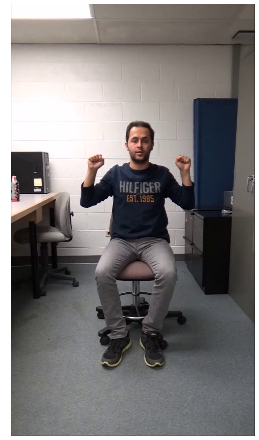
To be able to provide detailed feedbacks on each exercise, the first part of the exercise session is to perform only one repetition of the exercise. The eBear asks the user to say "*Start*" before starting, and "*Stop*" after finishing the one repetition: "*Once you say start I will show a start sign so that you know I heard you and I will show a stop sign as well.*". Once the user say start, the eBear shows a start sign on its screen to let the user know that the capturing is started. Stop sign is shown as well. The signs are shown in 5.9. The next step is to analyze this one repetition using the aforementioned method 2. We chose 70 out of



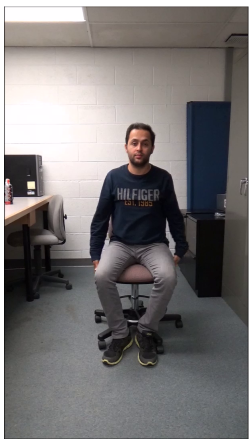
(a) Exercise 1 step 1

(b) Exercise 1 step 2

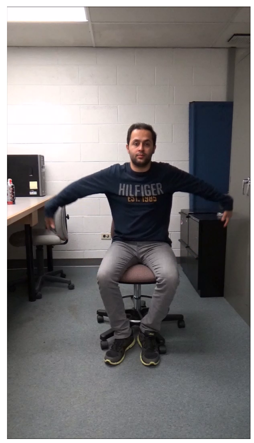
(c) Exercise 1 step 3



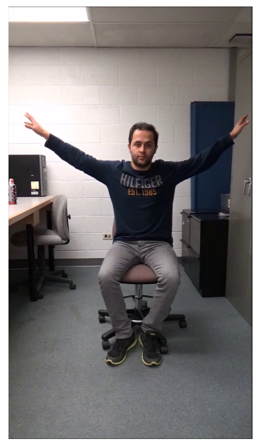
(d) Exercise 1 step 4



(e) Exercise 2 step 1



(f) Exercise 2 step 2

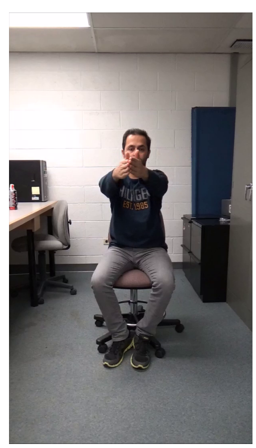


(g) Exercise 2 step 3

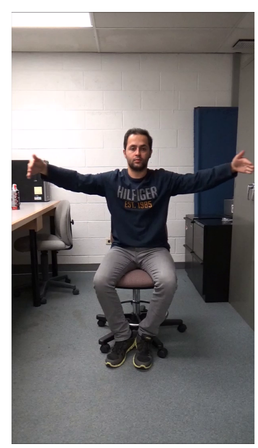
(h) Exercise 2 step 4

(i) Exercise 3 step 1

(j) Exercise 3 step 2



(k) Exercise 3 step 3



(l) Exercise 3 step 4

Figure 5.8: The major steps of the exercises

(a)

(b)

Figure 5.9: The Start and Stop signs used to show the starting and ending of each repetition of the first part of each exercise session.

100 as the threshold for the exercises. If the result was more than this threshold, the eBear proceeds to the next step, otherwise feedback is provided. According to the algorithm of method 2, if the gradient of each of the joints exceeds a threshold, feedbacks will be provided for that joint. Since the feedback is expressed in terms of numbers (direction in degree), we defined a set of sentences/commands for each joint. The direction is discretized into the main directions: up, down, left, right, back, forward. Depending on the direction, the eBear provides the related sentences/commands. For example, the sentences/commands for the right elbow are as follow:

- *You should move your right elbow up.*
- *You should move your right elbow down.*
- *You should move your right elbow left.*
- *You should move your right elbow right.*
- *You should move your right elbow back.*
- *You should move your right elbow forward.*

1 2 3 4 5

(a) (b) (c) (d) (e)

6 7 8 9 10

(f) (g) (h) (i) (j)

Figure 5.10: The numbers shown on the screen after the user counts after each repetition of the exercise

The feedback step is repeated until the score passes the threshold. There can be times when the user is not able to perform the exercise the same as the reference. After a few tries, if the user could not pass the score, the operator can pass the step manually.

In the next step, the user is asked to perform the same exercise 10 times. The eBear tells the user to say start at the beginning and count up after each repetition. Once each number is called the screen shows the same number (figure 5.10) so that the user makes sure that the eBear understood correctly.

After showing number 10, the eBear starts the analysis. Each repetition is analyzed separately and the average of the scores is reported to the user. If the average score is less than 70, the user is suggested to start the first step of the exercise again, otherwise the user can repeat the 10 times again, quit the exercise session, or change to another exercise.

We did not run mathematical evaluation on the method 2 of activity assessment. However it was integrated into the eBear which was tested in the field trial and the exit survey covered the exercise questions as well. The results are discussed in the evaluation chapter.

6

Field Test and Evaluation

This chapter presents the results of the comprehensive evaluation techniques which were conducted to evaluate the effectiveness of interacting with the eBear on health, mood, and participants perception of the robot and its characteristics. Each of the evaluation techniques focused on the validation of of a different aspect of the robot. Exposure to a social robot for the first time, might not provide an accurate evaluation of how the participants will perceive the robot in a period of a month. This is mainly due to the fact that interacting with an intelligent social agent is exciting and pleasurable since most of its features and characteristics are new to a person who has never been interacting with a social robot (31) (50). In order to evaluate different aspects of the eBear, a trial study was designed so that elderly people could interact with the eBear for a long period of time. The required Institutional Review Board (IRB) approvals were acquired for the human-subject involvements in the study. In the initial stages of the eBear development, it was taken to two of the assisted living houses in Denver, Eaton Senior Communities (15) and Mountain Vista Senior Living Community (4). Several elderly people at each of the centers interacted with the eBear in a Wizard-Of-Oz manner. At that time, they expressed their interest in interacting with such a platform. However, it was a Wizard-Of-Oz interaction and did not include any of the games or exercise sessions. For the final evaluation, several assisted living facilities in Denver of Colorado were asked for permission to take the eBear to their facility. Eaton Senior Communities (15) is one of the assisted living houses in Denver of Colorado which agreed to help with the study. They were contacted one month before the evaluation begins to start recruiting the participants. A set of flyers was distributed in the center which contained a general overview of the experiment. 7 Participants with none to severe depression were selected among the volunteers to participate in the study. Eaton center is a senior center and not all the members have depression and they do not run depression evaluation on the members. The selection of the participants were mainly based on the staff opinions on the depression levels of the residents. Table 6.1 illustrates the demographic and health information of the study participants. Throughout the study, in order to keep all the subjects information private, each participant was identified with an ID. As it is shown the subjects (n=7) ages range from 63 to 81 with average of 71.8($\pm$7.1). Each subject was supposed to interact with the eBear three times a week for a period of one month. Each session could take up to 1 hour. An introductory session was held with all the participants one week before the study. In the introductory session, each participants defined a schedule for the interaction sessions. A copy of the schedule was given to each person as their reminders. A consent form was also given to each of the participants. All the participants agreed to be recorded in the sessions. The library room was also scheduled to be available during the study.

6.1 The Study Setup

Figure 6.1 shows the setup that was used in the experiment. The eBear was placed in the library room of Eaton Communities Center. There were two couches in front of the eBear which subjects could sit on any of them. Three different lighting source as well as a two

Table 6.1: The participants information

Participant ID	Gender	Age	Condition
1	Female	64	
2	Female	63	
3	Female	78	Always on a scooter. Severe Depression.
4	Female	75	
5	Male	75	Mild Depression.
6	Male	81	Always with cane. Mild Depression.
7	Male	67	Mild Depression.

fans were used to provide a convenient place for the residents. The whole eBear software was installed on a Core-i7 laptop with 8GB of RAMs. As it is shown in Figrue 6.2, the laptop was put next to the eBear so that all the cables could be plugged in. The operator would sit in front of the laptop. With permission from all the subjects, all the sessions were recorded using the camera that was setup in figure 6.3. The participant, the eBear's face and screen, and the operator were in the sight of the camera.

Figure 6.1: Interaction Setup - The study setup of the the eBear

Figure 6.2: The Laptop - The laptop was put next to the eBear where the operator sit

Figure 6.3: A camera was setup to capture the sessions

6.2 Interaction Scenarios

Each participant interacted with the the eBear for three times per week for a period of a month in December of 2015. The session were held on Tuesday, Thursday and Saturday of each week. Only one operator was in the room for each session which were responsible for starting the eBear, capturing videos, and fixing any probable issues with the eBear. Each subject showed up at the designated time-sluts and sat on a couch in front of the eBear. Figure 6.4 depicts a participant sitting in front of the eBear ready for the interaction to start. Each session, the eBear started with greetings followed by a face scale quiz which described in section 3.2.3. The face scale quiz was performed at the end of each session as well. After the face scale quiz, they would continue the conversation with the eBear. In middle of the conversation, the eBear prompted the subject to start different programs that were integrated into the eBear in section 3.2. The subject could refuse to perform any of the programs at any time during the sessions. The eBear would prompt one or two of the programs randomly during each session to decrease the chance of being perceived as boring. The operator could also choose any of the programs at any time. Each resident was told to ask for help at anytime during each session when they needed.

6.3 Evaluation Methods

Comprehensive information were gathered throughout the study to assess the eBear and its effectiveness from different points of view. The resources include:

- Three different questionnaires: Geriatric Depression Scale, Almere Mode, and Exit Survey.
- All the sessions were recorded.

Figure 6.4: Example subject interacting with the eBear - Example participant sitting in front of the eBear ready for the interaction to start.

- The eBear logged all the conversations.
- The eBear logged all the facial expressions.
- The eBear logged all the face scale test results.

Below each of the evaluation methods is described along with the related results.

6.3.1 Exit Survey

An exit survey questionnaire was submitted in along with the consent form for IRB approval. The survey covers three main parts of the eBear: The general aspects of the eBear and its effects on their mood and wellbeing, the exercise program, and the games. All the questions were based on a five point Likert scale where 1 means strongly disagree and 5 means strongly agree with 3 being neutral. Table 6.2 illustrates the three parts of the questionnaire and the related questions.

A: Evaluation of the interaction with the robot	1. I enjoyed interacting with the robot
	2. Learning to interact with the robot was easy
	3. Talking with the robot was like talking to a person
	4. The robot was intelligent
	5. The robot was helpful
	6. The robot was acting natural
	7. I liked the robot's facial expressions
	8. The robot's facial expressions were natural
	9. I would like to have this robot at home
	10. I feel less depressed after talking to the robot
	11. The robot encouraged me to be more active
	12. The robot encouraged me to talk more
B: Evaluation of the exercise program	1. I enjoyed the exercise program

	2. The exercise program was entertaining
	3. I would rather have this robot as my exercise instructor rather than a real person
	4. I liked the way the robot tells the exercise instructions
	5. The exercise feedbacks were helpful
	6. The exercise feedbacks were accurate
C: Evaluation of the games	1. I enjoyed the games
	2. The games were entertaining
	3. I would rather have this robot as my game partner rather than a real person
	4. I liked the way the robot tells the game instructions.
	5. The games were helpful

Table 6.2: The questions of the exit survey

The participants were given the exit survey in the last session and they were given one week to fill out the survey. Figure 6.5 illustrates participants answers to the survey questions. The same answers are highlighted in the same colors.

Participant ID	A.1	A.2	A.3	A.4	A.5	A.6	A.7	A.8	A.9	A.10	A.11	A.12
1	5	4	4	4	4	3	5	4	5	5	5	4
4	5	5	3	4	4	3	5	4	4	5	2	5
2	5	5	5	5	5	5	5	5	5	5	5	5
3	5	5	4	4	3	3	5	4	4	5	3	4
6	5	5	4	4	4	4	4	4	3	4	4	4
5	4	5	3	3	3	4	3	3	2	3	3	4
7	3	2	3	3	2	3	4	3	2	4	4	3

(a) Part A

Participant ID	B.1	B.2	B.3	B.4	B.5	B.6
1	5	3	4	4	4	3
4	2	2	4	4	4	4
2	5	5	5	5	5	5
3	4	4	3	4	3	3
6	4	4	3	4	4	4
5	4	3	4	4	3	4
7	4	4	4	3	2	3

(b) Part B

Participant ID	C.1	C.2	C.3	C.4	C.5
1	4	4	4	4	4
4	5	5	4	5	5
2	4	4	4	4	4
3	5	5	3	4	4
6	4	4	3	4	4
5	3	4	3	3	3
7	4	3	3	3	4

(c) Part C

Figure 6.5: The answers to the exit survey

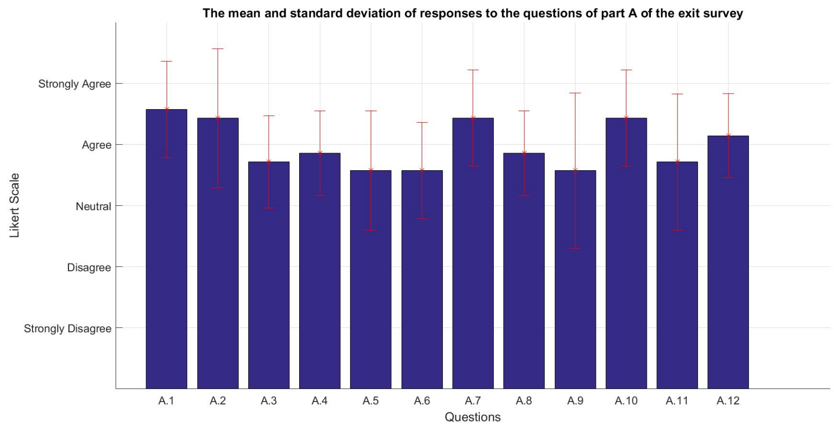
Figure 6.6 and figure 6.7 illustrate mean and median of the responses respectively. The participants enjoyed interacting with the eBear and learning to interact with the robot has been an easy task for them. Their impressions from the robot were not like talking to a real person. However on average rating is close to agree. The eBear is a animal-like robot which is designed to help people and the participants cannot be expected to perceive interaction with such a platform as interaction with a human. This result is also more understandable according to the fact that still there are many factors in a social agent which make the interaction it different than interacting with a real person. From a hardware point of view, for instance, in the eBear the movements of the mechanical servos were not natural in the sense that they do not have the smoothness that a human has while moving different limbs. Software wise, for instance, there were cases where the eBear misunderstood the speech which was probably because of the participants accents or environment noises. On average, most of subjects agreed that the eBear was an intelligent entity. This outcome can be justified by looking at the similarities between programs of the different interaction sessions. The purpose of the eBear is to assist people in a social manner which should be accompanied by an intelligent system but being intelligent was not the main goal of the eBear. The subjects almost agreed that interacting with the robot was helpful and the robot acted natural. The eBear's facial expressions were strongly liked by the subjects. This outcome was obvious even based on the subjects' comments on cuteness of the eBear's appearance and the facial expressions. The participants did not strongly agree to have the robot at home which can be justified by the fact that there was a operator next to the robot and the eBear needed some maintenance throughout the experiment. The target of this study was elderly people with depression and the results showed that the eBear could be a possible solution for treating depression as opposed to anti-depressed drugs. Most of the participants strongly agreed that they felt less depressed after interacting with the eBear. They roughly agreed that the eBear encouraged them to be more active which can be justified by the type of programs, only the exercise program would motivate them to be more active.

On average all the participants agreed on the part B of the questionnaire. This part was about the exercise program. Question 2 of this part had less ratings in compare to others which convey that the exercise programs were not entertaining enough. There were only three workouts designed for the eBear which were all seated exercises. The exercise program can be made more attractive by making it more interactive. This can be achieved by adding sound and more visual cues to the software. In the current version of the software, sound was not used in the exercise program.

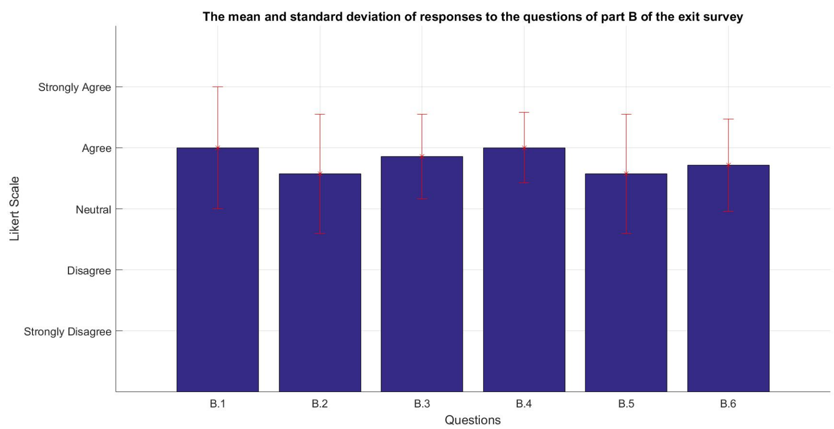
In part C, the same as part B, the subjects mostly agreed on all the questions. The question number three received the minimum rating. This question asked if they would prefer having the eBear as their gaming partner rather than a real person. This answer can be justified by looking at amount of conversation and excitment which is involved between the two sides of a game when both sides are human.

6.3.2 Geriatric Depression Scale

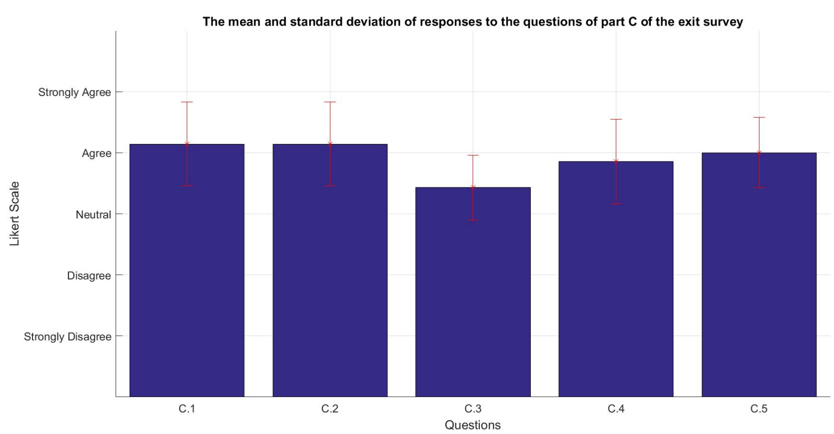
Geriatric Depression Scale or GDS is a screening tool used to identify depression in older adults. The general version of GDS contains 30 ”yes” or ”no” questions. There is a short version of GDS or GDS-SF which only has 15 questions. GDS-SF has been proved to be a adequate replacement for the original scale (57). Figure 6.8 depicts the questionnaire of the short version of Geriatric Depression Scale which was used in this thesis. Each question is either 0 or 1 point and there are 15 points in total. The depression is identified as following:



(a) Part A

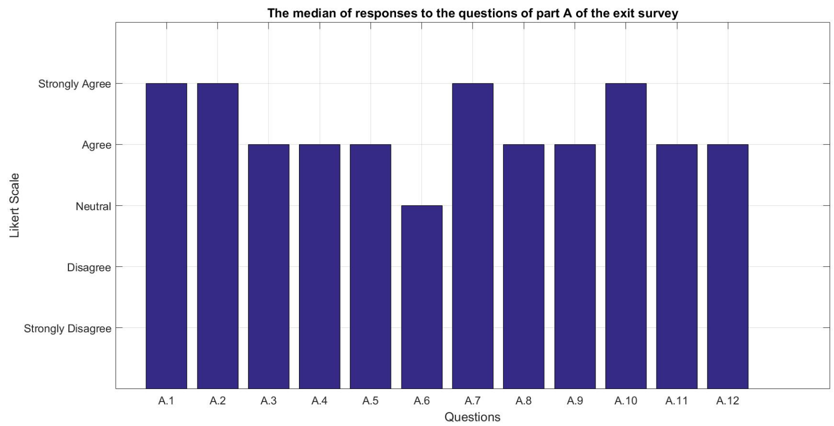


(b) Part B

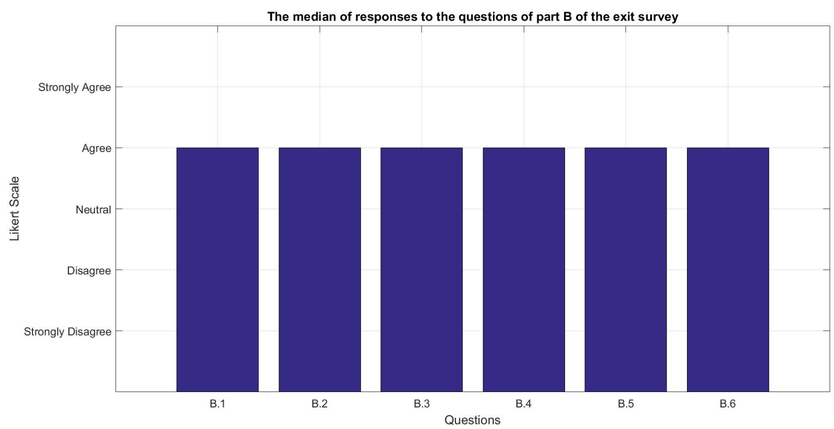


(c) Part C

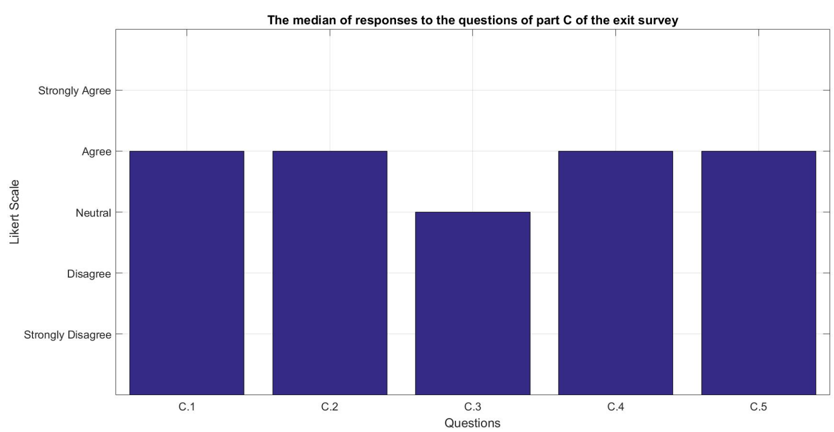
Figure 6.6: The mean of the responses to the exit survey



(a) Part A



(b) Part B



(c) Part C

Figure 6.7: The median of the responses to the exit survey

- Between 0 and 5 is normal
- A score more than 5 suggests depression
- A score more than 10 suggests severe depression.

Although GDS has been proven to be a reliable scale, but usually it is used along with other comprehensive geriatric assessments. In this thesis GDS is used along with several other measures, in particular the staff comments on the participants mood. The elderly individuals of this study interacted with the eBear for a month. In the beginning of each week they filled out one GDS form.

Figure 6.9 illustrates the measured GDS over the period of one month for each of the subjects. As it can be seen, in the beginning of the study, participants 3 and 6 had GDS scales of more than 5 which suggest depression. The nursing staff comments on these two participants strengthened the possibility of depression among them. Participant number 5 started with a scale of 5 which is the limit between two areas. Participant number 1 had a score of 0 and the rest of the participants' scores were between 0 and 5.

Subject 1 started with a scale of 0 and stayed the same in all the four measures. She was very active and the caregivers confirm the subject mood. Participant number 5 started with a score of 5 and the scale constantly dropped during the course of the study down to scale of 2. Participant number 4 showed almost a similar pattern to participant 5. However the scale started from 2 and ended in 1, hence the differences were not the same between the two subjects. These two GDS patterns are fully along with the objective of this study. Scales of subjects 2 and 7 had some fluctuations during the experiment but stayed close to 0 in all the four weeks. Subject 3 was a 78 year old elderly individual with mild depression. According to the caregivers, she had been feeling down due to severe health and emotional problems. As the bar chart shows, she started with a scale of 7 which incremented to 8 in the next after. However, the scale dropped to 6 and 4 in the next two weeks. This trend could be easily seen in her behavior which surprised the caregivers. Subject 3 started with a scale of 6 but dropped down below 5 in the following weeks. The GDS scale of subject 3 dropped down by 3 scales. This trend is along the goal of this thesis.

In conclusion, the GDS scores of this study backed the hypothesis that a social robot such as the eBear can have uplifting effects on the elderly individuals mood.

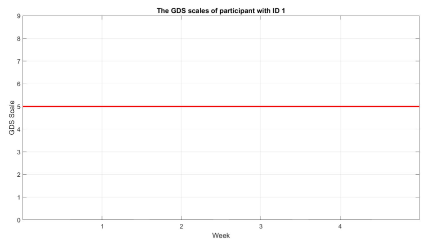
***Instructions**: Choose the best answer for how you felt over the past week.*

No.	Question	Answer	Score
1.	Are you basically satisfied with your life?	YES / NO	
2.	Have you dropped many of your activities and interests?	YES / NO	
3.	Do you feel that your life is empty?	YES / NO	
4.	Do you often get bored?	YES / NO	
5.	Are you in good spirits most of the time?	YES / NO	
6.	Are you afraid that something bad is going to happen to you?	YES / NO	
7.	Do you feel happy most of the time?	YES / NO	
8.	Do you often feel helpless?	YES / NO	
9.	Do you prefer to stay at home, rather than going out and doing new things?	YES / NO	
10.	Do you feel you have more problems with memory than most people?	YES / NO	
11.	Do you think it is wonderful to be alive?	YES / NO	
12.	Do you feel pretty worthless the way you are now?	YES / NO	
13.	Do you feel full of energy?	YES / NO	
14.	Do you feel that your situation is hopeless?	YES / NO	
15.	Do you think that most people are better off than you are?	YES / NO	
		TOTAL	

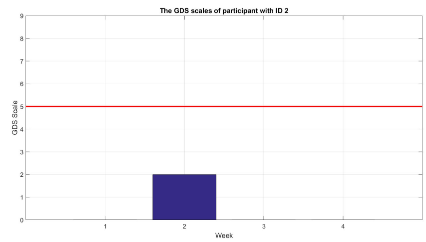
Figure 6.8: Geriatric Depression Scale - The questionnaire of short version of Geriatric Depression Scale

6.3.3 The Face Scale

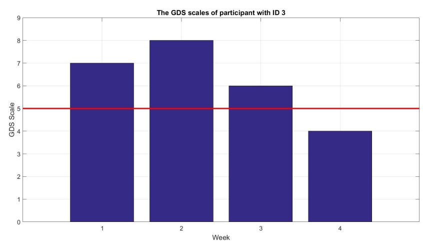
The Face Scale mood evaluation method was described in 3.2.3. Basically it is consisted of a set of genderless face pictures with different moods. The patient is supposed to pick the face with the mood that better matches his/her feelings. A software was designed and



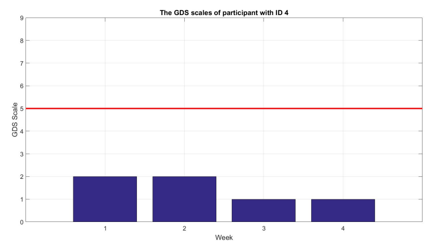
(a) Participant 1



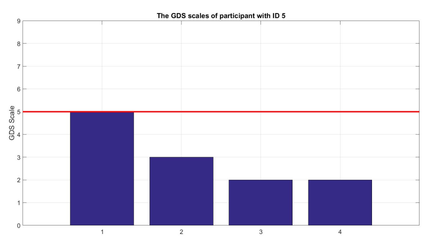
(b) Participant 2



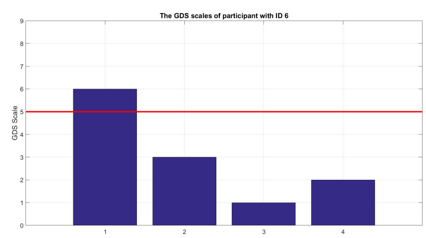
(c) Participant 3



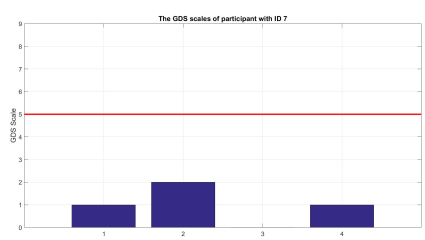
(d) Participant 4



(e) Participant 5



(f) Participant 6

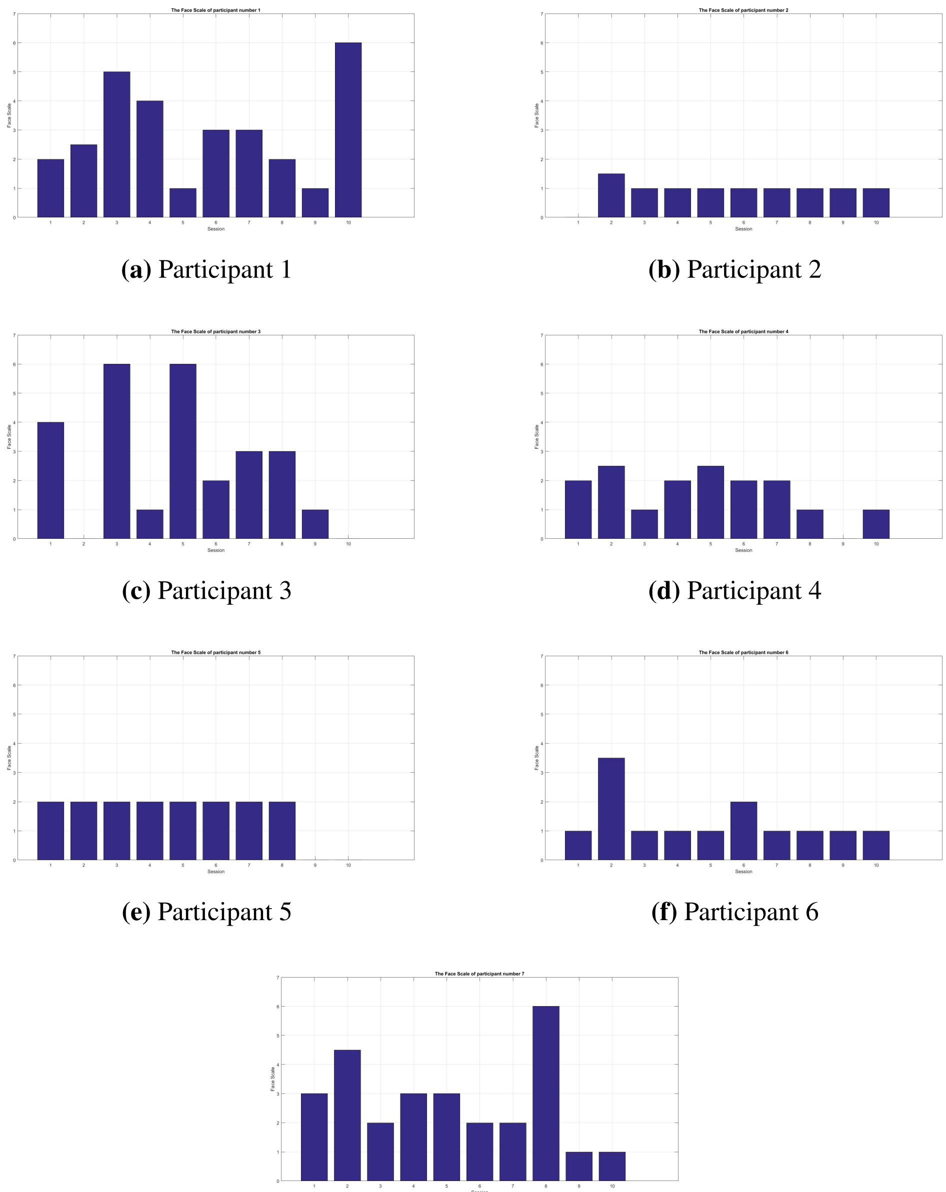


(g) Participant 7

Figure 6.9: The GDS scale of each week

developed so that the eBear could automatically run the Face Scale method. In the Face Scale program the user is asked whether he/she wants to tell the eBear about his/her current mood. In each session, the eBear prompts the Face Scale program two times, once at the beginning of the session and once at the end of the session. Figure 6.10 illustrates the Face Scale results of each participant. In the charts, some of the points are missing which is either because they refused to participate or because they did not show up in that session. The mean and the related standard deviation, and median of the results are depicted in Figure 6.11.

According to the average and median of the results, the general mood of the participants improved over the course of the experiment. However, the improvement is minor and happened mostly in the last two sessions.The result of each participant better justifies the patterns of the mean and the median. Participant 1 started with a scale of 2 which is considered as a good mood. Her Face Scale score increased to 5 but decreased afterward except the last session when rocketed up to 6 which is the worst case. In the mean chart, the standard deviation of the last bar shows this behavior. Participant 1 did not have depression which can be seen from Figure 6.9. In the last session she was not in mood and her behavior in the last session is less likely to be related to the eBear and might had been affected by an external cue such as family problem. The scores of subjects 2 and 5 stayed almost the same around scores 1 and 2, respectively. These scores show a good mood. The score 1 of participant 2 is along her GDS scores which were almost zero. However, for subject 5 the GDS scores decreased while the Face Scale score stayed almost the same. The general pattern of the Face Scale scores of subject 3 matches her GDS scores pattern, although with more fluctuations. The scores suggest that subject 3 mood improved over the experiment. Participants 4 and 6 mood scores fluctuated between 1 and 3.5 and do not follow a specific pattern. For subject 4, the GDS scores did not have much variations. However, as for subject 6, the GDS scores improved significantly over the time. Subject 7 scores showed



(a) Participant 1

(b) Participant 2

(c) Participant 3

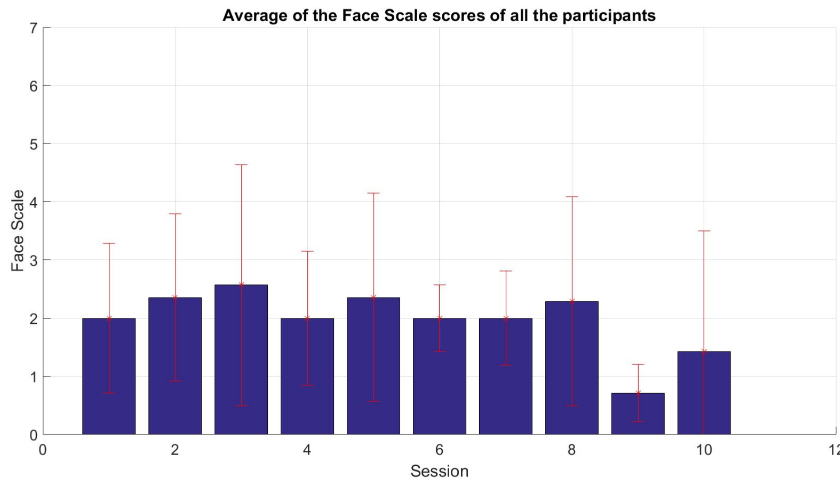
(d) Participant 4

(e) Participant 5

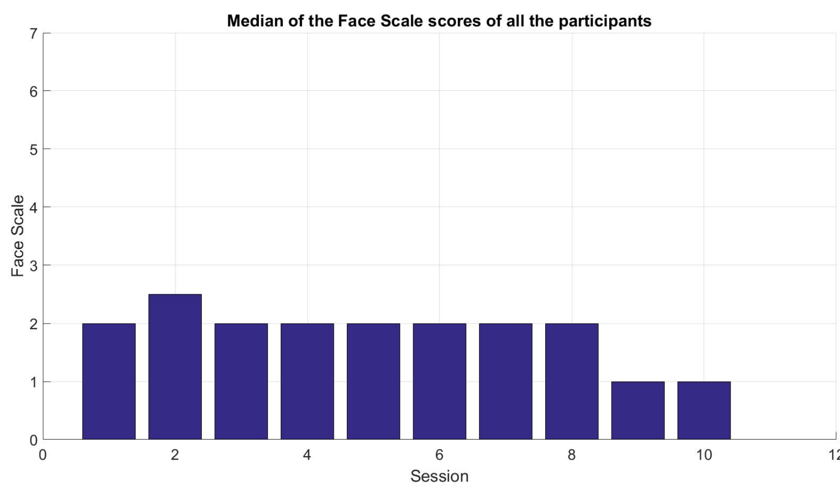
(f) Participant 6

(g) Participant 7

Figure 6.10: The Face Scale results



(a) Average of the results



(b) Median of the results

Figure 6.11: Average and Median of the Face Scale results

some irregular pattern where in two of the sessions the mood was bad, but for the rest the mood improved smoothly. In general, the results of the Face Scale tests followed almost the same pattern of the GDS scores. They proved that interacting with the eBear can have positive effects on people mood. However, the results suggest that using this evaluation method with only 6 mood scales might not be a precise method for mood evaluation.

6.3.4 Almere Model

Almere Model is a technology acceptance model designed to assess the acceptance of social robots by elderly people (45). Almere Model is designed to not only take into account the functional evaluations of the robot such as usefulness, but also the social interaction itself which is one important objective of this thesis. Table 6.3 depicts the questionnaire of this model. As it is shown, each question is related to a code. The codes are defined in Table 6.4 and each code is related to a construct. Basically each context evaluates a specific aspect of the robot. The same as the GDS, the participants filled out this questionnaire at the beginning of each week. Each question should be scored based on a Likert scale from 1 to 5: 5=strongly agree, 4=somewhat agree, 3=neutral, 2=somewhat disagree, 1=strongly disagree. Figure 6.12 illustrates the average of the responds to each question among all the participants. In Almere Model, usually the results are expressed as the average of each construct. Figure 6.13 depicts the average of each construct and the standard deviations. The other measure which is usually used in Almere Model, and in general in Likert scales, is the Cronbach's Alpha measure (76). This measure expresses how closely related a set of values are as a group which can be viewed as a reliability statistics. When the alpha value is higher, it suggests that the data have relatively high internal consistency. Figure 6.14 shows the alpha value for each of the constructs. In order to better analyze the subjects data, the results for each participant is separately depicted in Figure 6.15.

The lowest score in Figure 6.13 is for *ANX* construct. *ANX* stands for anxiety and represents the emotional or anxious reactions toward the eBear. In fact, *ANX* is the only negative construct and the less score means people feel less anxious when interacting with the robot. The alpha coefficient of this construct is 0.52 which does not strongly suggest that this construct has a high internal consistency. The results of each question reveals the reason of the medium alpha coefficient. As it is shown, the scores of this construct's questions vary from 1.36 to 2.61. The score 2.61 is for *ANX3* which asks if the user is afraid to make mistakes with it. This result can be justified by the fact that even though the robot was introduced in the introductory session, still this was their first experience of interacting with an automatic robot. The user-specific results of *ANX* construct shows that participant 7 gave the highest score of 3.72 to this construct. In the first few sessions, participant 7 expressed his anxiety to the operator. During the course of the study he was explained a few times that there is no need to always say yes to the questions that the eBear asks. The other question which carries a negative meaning is *PENJ5* which asks the user if the robot was boring. This question has a score of 1.5 which means the participants did not find the robot boring. An alpha coefficient of 0.72 suggests that most of the participants agreed on th point that the eBear is not boring. This behaviour can be seen in the results of each participants in Figure 6.15. *PAD* construct has the only negative value in the Cronbach's alpha chart. *PAD* evaluates if the robot is adaptive to the need of the user. The score of 3.63 suggests that in general participants thought that the eBear is adaptive to their needs. However, the scores varies from 3.51 to 4.38 which shows rather large variety. The other large variety is in *FC* construct which varies from 3.74 to 4.86 that is a high value.

Code	**Questions**

ANX	1. If I should use the robot, I would be afraid to make mistakes with it 2. If I should use the robot, I would be afraid to break something 3. I find the robot scary 4. I find the robot intimidating
ATT	1. I think its a good idea to use the robot 2. The robot would make life more interesting 3. Its good to make use of the robot
FC	1. I have everything I need to use the robot 2. I know enough of the robot to make good use of it
ITU	1. I think Ill use the robot during the next few days 2. Im certain to use the robot during the next few days 3. I plan to use the robot during the next few days
PAD	1. I think the robot can be adaptive to what I need 2. I think the robot will only do what I need at that particular moment 3. I think the robot will help me when I consider it to be necessary
PENJ	1. I enjoy the robot talking to me 2. I enjoy doing things with the robot 3. I find the robot enjoyable 4. I find the robot fascinating 5. I find the robot boring
PEOU	1. I think I will know quickly how to use the robot 2. I find the robot easy to use 3. I think I can use the robot without any help 4. I think I can use the robot when there is someone around to help me 5. I think I can use the robot when I have a good manual

PS	1. I consider the robot a pleasant conversational partner 2. I find the robot pleasant to interact with 3. I feel the robot understands me 4. I think the robot is nice
PU	1. I think the robot is useful to me 2. It would be convenient for me to have the robot 3. I think the robot can help me with many things
SI	1. I think the staff would like me using the robot 2. I think it would give a good impression if I should use the robot
SP	1. When interacting with the robot I felt like Im talking to a real person 2. It sometimes felt as if the robot was really looking at me 3. I can imagine the robot to be a living creature 4. I often think the robot is not a real person 5. Sometimes the robot seems to have real feelings
Trust	1. I would trust the robot if it gave me advice 2. I would follow the advice the robot gives me

Table 6.3: Almere Model questionnaire (45)

Code	Construct	Definition
ANX	Anxiety	Evoking anxious or emotional reactions when it comes to using the system
ATT	Attitude Towards Technology	Positive or negative feelings about the appliance of the technology

FC	Facilitating Conditions	Factors in the environment that facilitate use of the system
ITU	Intention to Use	The intention to use the system over a longer period in time
PAD	Perceived adaptiveness	The perceived ability of the system to adapt to the needs of the user
PENJ	Perceived Enjoyment	Feelings of joy/pleasure associated with the use of the system
PEOU	Perceived Ease of Use	The degree to which one 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 the system would be assistive
SI	Social Influence	The persons perception that people who are important to him think he should or should not use 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

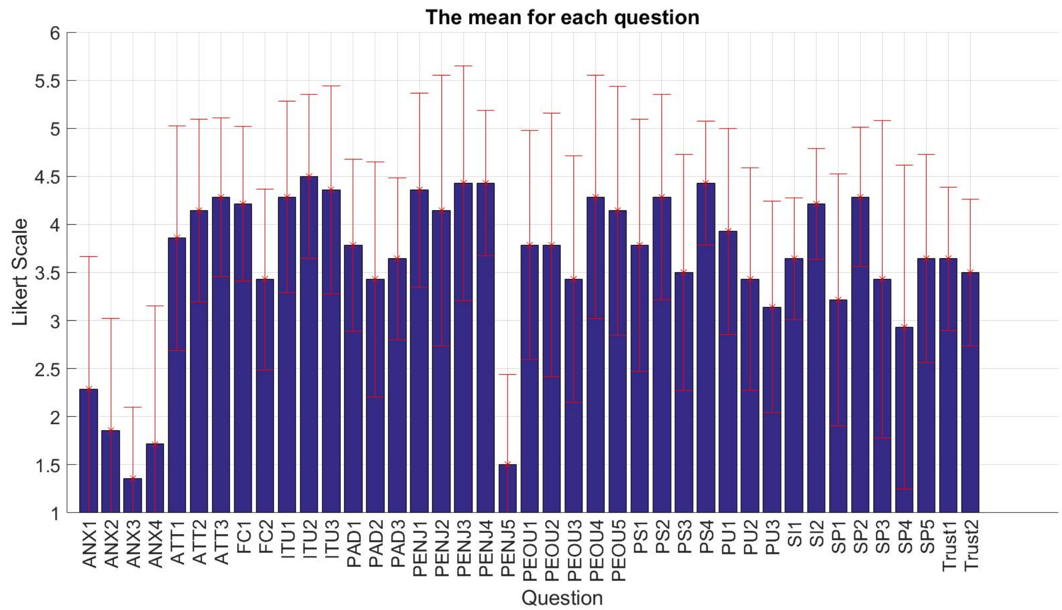


Figure 6.12: The mean of the results of each question of Almere Model for all the participants

Use	Use	The actual use of the system over a longer period in time

Table 6.4: Almere Model constructs and their definitions (45)

6.3.5 Observation Sheet

In order to better analyze the behavior of the subjects while interacting with the eBear, all the sessions' videos were charted on a minute-basis according to an observation sheet (89). A member of our team charted all the interaction videos based on the Table 6.5. This table illustrates the different items which were analyzed in this evaluation method. As it is shown, four classes are included in the observation sheet. Class "Context" is used to specify which program was mostly run in that particular minute of the video. In the "Expression"

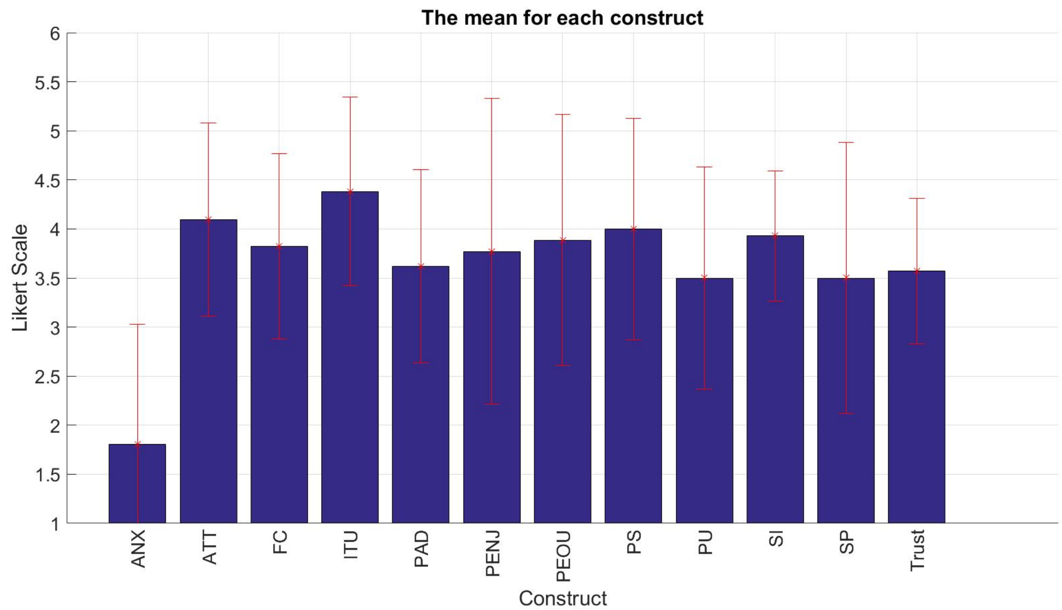


Figure 6.13: The mean of the results of each construct of Almere Model for all the participants

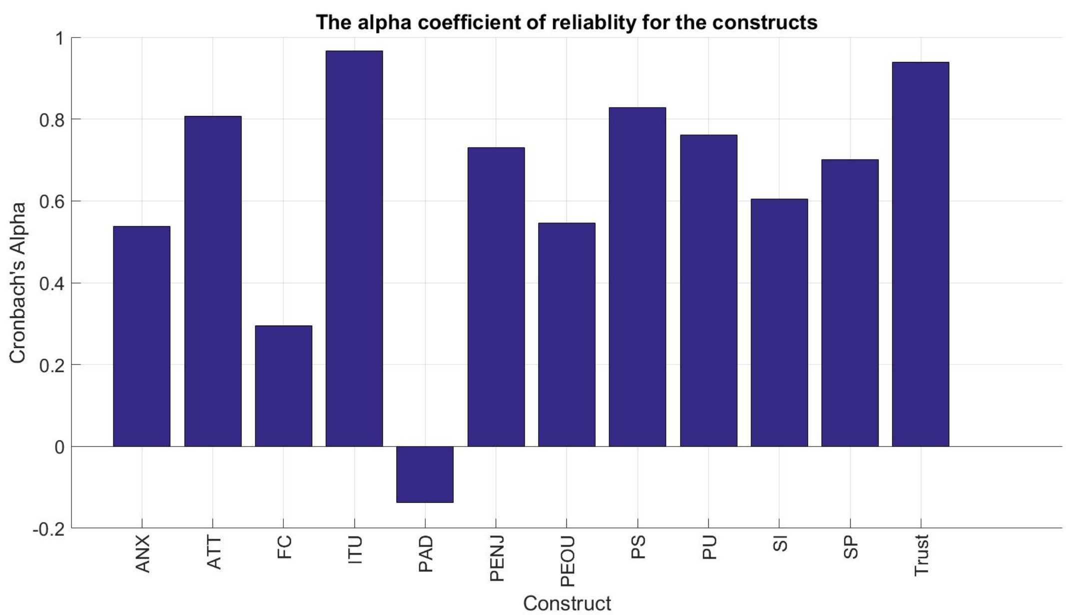
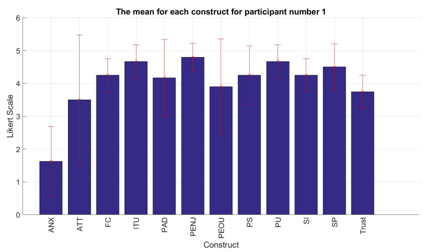
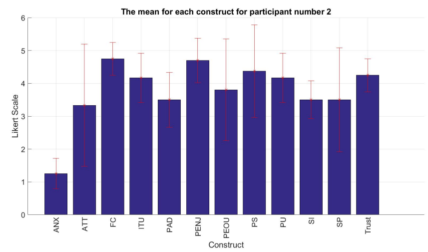


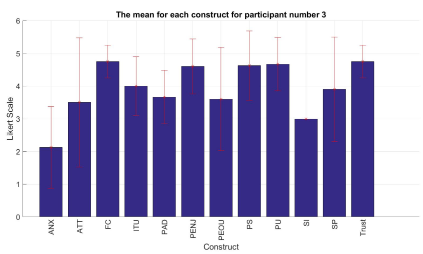
Figure 6.14: The Cronbach's Alpha value for each of the constructs



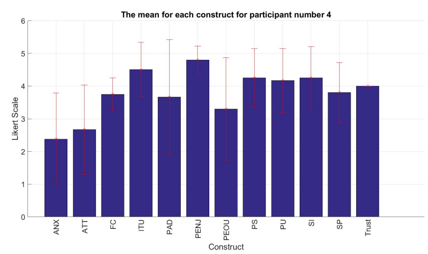
(a) Participant 1



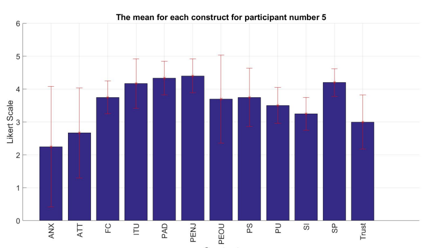
(b) Participant 2



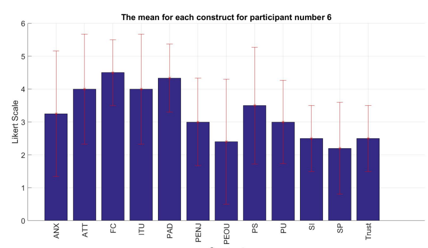
(c) Participant 3



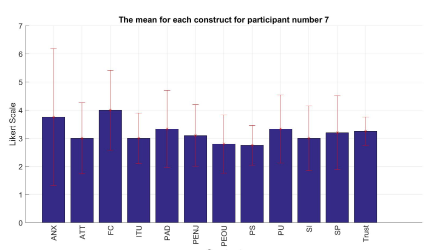
(d) Participant 4



(e) Participant 5



(f) Participant 6



(g) Participant 7

Figure 6.15: The mean of Almere Model results for each participant

class, the dominant expression of the subject is recorded. Class ”Gaze” specifies in which direction the subject was looking at in that minute.

We were interested in evaluating the effect of each context on the subjects mood. Table 6.6 illustrates the overall percentage of having each facial expression in different context of the interaction scenarios. Laughter and smile were most frequently observed while talking to the robot. The order of the results in ”Laugh” and ”Smile” expressions are totally the same. After ”Conversation” , playing games and watching videos were the most effective ones in bringing smile and laugh to the subjects faces. The subjects had the least amount of laugh and smile in the mood evaluation methods. The evaluation methods showed the same set of pictures all the times and it might be a possible reason for not smiling or laughing while answering the mood question. Another justification is that the nature of the mood evaluation method is in a way that does not make people smile or laugh. This is proved by the percentage of ”Neutral” expression which is more than all the other expressions. By neglecting three 0.09% cases, there were not any ”Sad” or ”Hate” minutes except in all the interaction scenarios. The ”Surprise” expression was most frequently observed while watching videos. Table 6.7 illustrates the overall percentage of having each facial expression versus the direction at which the user was looking at. Table 6.7 shows the overall percentage of having each facial expression versus the speaker.

Table 6.5: Code of observation sheet

Class	Item	Code
Context	Conversation	C
	Video	V
	Game	G
	Evaluation	Ev
	Others	O
Expression	Laugh	L
	Smile	S
	Sad	Sa
	Hate	H
	Surprise	Su
	Neutral	N
Gaze	eBear	E
	Screen	S
	Caregiver	C
	Others	O
Talk	eBear	E
	Participant	P
	Caregiver	C

Table 6.6: The overall percentage of each facial expression with respect to each context of the interaction scenarios

Context \ Facial Expression	**Laugh**	**Smile**	**Sad**	**Hate**	**Surprise**	**Neutral**
Conversation	5.98	15.11	0.09	0	0.47	2.56
Exercise	0.85	5.70	0	0.09	0.28	10.74
Video	2.18	5.98	0.09	0	0.76	5.70
Game	2.94	9.31	0	0	0.09	15.20
Evaluation	0.28	3.70	0	0	0.09	5.79
Others	0.19	1.61	0	0	0.28	3.80

Table 6.7: The overall percentage of each facial expression with respect to the subject's gaze

Gaze \ Facial Expression	Laugh	Smile	Sad	Hate	Surprise	Neutral
eBear	6.54	18.76	0.09	0	0.67	3.94
Screen	5.38	20.50	0.09	0.09	0.86	35.70
Operator	0.57	2.30	0	0	0.48	3.94

Table 6.8: The overall percentage of each facial expression with respect to the speaker

Speaker \ Facial Expression	Laugh	Smile	Sad	Hate	Surprise	Neutral
eBear	3.32	14.82	0	0	1.23	29.18
Participant	8.46	24.42	0.19	0.09	0.38	9.03
Operator	0.66	2.18	0	0	0.38	5.60

7

Conclusion and Future Research Direction

7.1 Conclusion

The purpose of this thesis was to present the design and development of the eBear, an expressive animal-like robot equipped with artificial intelligence to socially assist elderly people, in particular those with depression. Being ”Semi-Autonomous”, in the sense that the eBear runs all the modules automatically but the operator should schedule the different programs, and ”Proactive Engagement” are the terms which made the eBear unique. The eBear successfully interacted with 7 elderly people with depression for a month in an automatic manner. To the best of our knowledge, the eBear is the first social robot containing all of the following features in one platform:

- The eBear is semi-autonomous in the sense that all the modules are executed automatically, but the order and the start time of each the module was determined by the operator.

- The eBear could be used in a home or a care-center setting without the need for any special setup.

- The eBear could analyze the emotions of users from both visual and speech cues at the same time.

- The eBear could automatically run a mood evaluation method which could be used to assess the mood remotely.

- The eBear is proactive in prompting and motivating the user to involve in the different programs.

- The eBear could engage and carry out an open conversation.

- The eBear could motivate and coach users through exercise and provide them with the required feedbacks.

- The eBear could play different games with the user.

Chapter 6 presented the evaluation methods of this thesis. The results showed that the eBear could make improvements on elderly people lives. With the rising percentage of the aged population, such an automatic platform could play a major role in providing high quality individualized health-care in the near future. As discussed before, the purpose of such platforms is not to remove the need for a care-giver and to act as an assistant to the caregiver. One item which was not discussed in the thesis evaluation is the comments of the caregivers. As of the time of this thesis, due to the lack of enough time, only two of the staff members of Eaton Community Center were asked to report on the changes in the

Hello from Eaton,

I just wanted to thank you for allowing our residents the opportunity to experience E-Bear. The residents are enjoying helping Amir complete his studies. Amir has been wonderful with our residents. He has a kindness about him and is working hard to make this experience positive for the attendees. Tonight Amir reported unexpected outcome with one resident. [redacted]. has struggled with severe depression due to her health failing her. She went to her session saddened but came out with the biggest grin on her face and laughter! This was awesome since she is dear to so many of us. Thank you for that moment!
Nine News happened to be here to do a video clip on another student. The residents and the student were telling him about his robot bear and he is going to speak to someone at Nine News about a video segment with Amir at Eaton. He will be sending me contact information and I will pass that on to you for consideration. I expect Eaton to be on board with this if you are. Thank you again, you have no idea how this will impact future seniors!

Figure 7.1: A surprising email from Eaton Senior Communities

participants mood. Participant 3 had severe depression and according to the staff, she had not even laugh for a while before the eBear study started. It was a thrilling moment when her caregiver contacted the eBear's team to tell how happy subject 3 was after interacting with the eBear. Figure 7.1 illustrates this email. This mood change was clear in the evaluation results, however the comments of the caregivers showed how significant the change was on the subject 3 mood. Another case is subject 5 who had mild depression. According to the caregivers' comments, subject 5 asked for the eBear after the study was finished. He told the caregivers about his experience with the eBear, in particular the games and it seems like he enjoyed to interact with such a companion-bot on a daily basis. The results of this study even caught the attention of several other organizations in Denver Colorado which ended up in a media coverage from 9NEWS, a local TV. Figure 6.14 shows the 9NEWS interview with two of the subjects who participated in the study. These promising results can be viewed as a motivation for the researchers in this area to develop more adaptive social robots.

This study was based on a single-subject evaluation method. Basically, each subject was monitored throughout the study to see whether the eBear would improve their mood and wellbeing or no. Another type for evaluation is to use two groups, one study group and one control group. Usually these two groups are chosen to be similar in terms of their

Figure 7.2: 9NEWS interview with several of the subjects

condition (e.g. depression level). The study group would interact with the robot according to their schedule, but the control group will be just monitored and evaluated as a basis for the evaluation results of the study group. In the case of the eBear, the results of the GDS, Almere Model, or the Face Scale could be better analyzed if compared to the results of a control group. In order to further analyze different features of the eBear, some features could be deactivated for some subjects while the rest of the subjects are using those features. These features can be the visual and speech emotion recognition modules. In this study, it was observed that the subjects enjoyed the facial expression mirroring feature of the eBear.

The process of design and development of the eBear had various invaluable lessons which not necessarily a computer engineering student comes across. The eBear development consisted of two major parts: hardware and software. In order to make an automatic robotic platform, each of these parts and their subparts should meet at least some minimum requirements. While designing the facial expressions of the eBear, we learned how to design the appearance of a robotic platform in a way that people perceive it as likable and desirable. The mechanical modifications of the eBear such as the 3D printed tablet frame thought us how to design a basic 3D object in Solidwork software and all the processes involved in 3D printing. Since the initial stages of the planning for the eBear, we studied various psychological research publications, consulted with several psychologists, and met with a few senior community centers and their elderly residents to have a better understanding of the target population of the eBear. Meeting with the elderly people with depression and getting to know their needs, gave us significant motivations while designing the eBear.

Designing software and hardware took time and effort. However, hardware designing and development introduced many challenges which not necessarily carry a scientific value. For example, there is a tablet integrated into the eBear's body which is used as a screen. It took more than a week to design and 3D print a frame for the tablet, cut and weld an outer frame inside the eBear, whereas integrating a screen to a robot might seem to be an easy task which definitely does not contain any scientific contribution.

As of the time of this thesis, the research of the eBear platform is not published yet. Two research papers were published as following.

- Kargar, B.A.H.; Mollahosseini, A.; Struemph, T.; Pace, W.; Nielsen, R.D.; Mahoor, M.H., "Automatic measurement of physical mobility in Get-Up-and-Go Test using kinect sensor," inEngineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE, vol., no., pp.3492-3495, 26-30 Aug. 2014

- Xiao Zhang; Mollahosseini, A.; Kargar B, A.H.; Boucher, E.; Voyles, R.M.; Nielsen, R.; Mahoor, M.H., "eBear: An expressive Bear-Like robot," inRobot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on, vol., no., pp.969-974, 25-29 Aug. 2014

In general, according to the evaluation methods used in this thesis as well as the Eaton Senior Community staff members comments, the eBear is proved to be an effective companion robot capable of uplifting mood of the elderly people with depression. Because the duration of the trial study was one month and after the period, the residents showed interest in interacting with the eBear again, this proves that the eBear is not a boring social platform.

7.2 Future Work

The eBear was designed and developed with the ultimate goal of deployment in the house of elderly individuals with need. In order to build a robot to be used by an end user who might not have any familiarity with technology, first we will need to make the eBear fully autonomous. In this way, it can be left at the house of the elderly people. In this study people interacted with the eBear for a period of one month, however a longer period of interaction might reveal more advantages and disadvantages of the eBear which were not clear in this study. The other potential future research direction is to change the target population. For example, the elderly people with dementia or children with Autism also might benefit from such a robotic platform. During this study, all the interactions were based on a one-on-one communication, one should change conduct the same study with a group of people in front of the eBear.

One issue which was observed during this study is the fact that the current speech recognition systems do not correctly recognize all voices. In same cases the speech recognition modules failed to correctly recognize the speech. Since the target population is elderly people, we believe there should be more research toward using as many of the human senses as possible in the interactions. For example, if the robot failed to understand the user's speech, the user inputs the required information using the touch screen. Here we describe the rest of the items which could be added in the next iteration of the eBear:

Hardware

- The eBear should have a more robust and more durable mechanical design for the head. Users even might hug the robot and the robot's body should be able to tolerate such forces.
- The eBear's servos make noise while they are being used. These noises could be annoying in a home setting where is usually quite, in particular in case of an elderly user. The servos should be replaced with stronger servos to avoid the noise.
- The eBear's screen is a 10 inch display which is small for elderly users, particularly the ones with weak eyes. Another solution can be adding a remote screen or even connecting to the house TV.
- Integrating mechanical arms to the eBear adds more functionality which results in more interactive programs.
- The servos, sensors of the eBear heats up while they are in use. A ventilation system should be added to cool down the platform.
- The eBear needs the Internet for several of its functionalities. A GSM cellular communication device should be utilized so that the eBear could be freely used in all the places.

Software

- The eBear used UDP to communicate between the screen and the main laptop. Better communication protocols should be used in order to eliminate the risk of loosing data packets.

- The open conversation of the eBear was done via an online chatter-bot service. A private natural language process module should be developed for the open conversation module so that it could be modified based on the mood of the user and the situation.

- The games of the eBear were not attractive enough. The eBear should be more engaged while playing the games. The engagement should be in a way that the user feels that he/she is playing with the eBear and not only on the eBear's screen.

- A protocol should be designed so that th eBear sends all the acquired information to the caregiver.

- A module should be designed so that the eBear could be also be used in a Wizard-Of-Oz manner.

- The eBear should be able to act as a reminder for different daily schedules of the user.

- The eBear should be able to make phone calls to emergency contacts of the user.

With these modifications to the eBear's platform, it will be one step closer to the final goal of deploying in the house of the user.

References

[1] http://aylien.com/. 40

[2] http://hitecrcd.com/. 23

[3] http://jpn.nec.com/robot/en/. 13

[4] http://mountainvista.net/. 83

[5] http://msdn.microsoft.com/en-us/library/hh973074.aspx. 49, 54, 59, 60, 75

[6] http://neospeech.com/. 41

[7] http://robotics.usc.edu/. 12

[8] https://software.intel.com/en-us/intel-realsense-sdk. 39

[9] https://www.census.gov/. 1

[10] https://www.pololu.com/. 23

[11] http://wongbakerfaces.org/. ix, 44

[12] http://www.bartneck.de/. 17

[13] http://www.bavieca.org/. 23

[14] http://www.cleverbot.com/. 38

[15] http://www.eatonsenior.org/. 83, 84

[16] http://www.hagisonic.com. 19

[17] http://www.pandorabots.com/. 38, 40

[18] http://www.parorobots.com/. 8

[19] http://www.research.philips.com/. 16

[20] http://www.sony-aibo.com/. 10

[21] American Psychiatric Association et al. The diagnostic and statistical manual of mental disorders: Dsm 5. bookpointus, 2013. 2

[22] Marian R. Banks, Lisa M. Willoughby, and William A. Banks. Animal-assisted therapy and loneliness in nursing homes: Use of robotic versus living dogs. Journal of the American Medical Directors Association, 9(3):173 – 177, 2008. 11

[23] Elizabeth E Baum, David Jarjoura, Ann E Polen, David Faur, and Gregory Rutecki. Effectiveness of a group exercise program in a long-term care facility: a randomized pilot trial. Journal of the American Medical Directors Association, 4(2):74–80, 2003. 2, 58

[24] Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In KDD workshop, volume 10, pages 359–370. Seattle, WA, 1994. 65

[25] Timothy W Bickmore and Rosalind W Picard. Establishing and maintaining long-term human-computer relationships. ACM Transactions on Computer-Human Interaction (TOCHI), 12(2):293–327, 2005. 3, 7

[26] AS Billis, EI Konstantinidis, C Mouzakidis, MN Tsolaki, C Pappas, and PD Bamidis. A game-like interface for training seniors dynamic balance and coordination. In XII

Mediterranean Conference on Medical and Biological Engineering and Computing 2010, pages 691–694. Springer, 2010. 4, 7

[27] KM Culhane, M O'Connor, D Lyons, and GM Lyons. Accelerometers in rehabilitation medicine for older adults. Age and ageing, 34(6):556–560, 2005. 45

[28] Charles Darwin. The expression of the emotions in man and animals, volume 526. University of Chicago press, 1965. 30, 32

[29] Doreen Dawe and Robin Moore-Orr. Low-intensity, range-of-motion exercise: invaluable nursing care for elderly patients. Journal of advanced nursing, 21(4):675–681, 1995. 58

[30] George Demiris, Marilyn J Rantz, Myra A Aud, Karen D Marek, Harry W Tyrer, Marjorie Skubic, and Ali A Hussam. Older adults' attitudes towards and perceptions of'smart home'technologies: a pilot study. Informatics for Health and Social Care, 29(2):87–94, 2004. 45

[31] Carl F DiSalvo, Francine Gemperle, Jodi Forlizzi, and Sara Kiesler. All robots are not created equal: the design and perception of humanoid robot heads. In Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques, pages 321–326. ACM, 2002. 83

[32] Paul Ekman and Wallace Friesen. Facial action coding system: A technique for the measurement of facial movements. Consulting Psychologist, 2, 1978. 30

[33] Juan Fasola and Maja J Matarić. Robot exercise instructor: a socially assistive robot system to monitor and encourage physical exercise for the elderly. In RO-MAN, 2010 IEEE, pages 416–421. IEEE, 2010. 7, 13, 67

[34] Li Fei-Fei and Pietro Perona. A bayesian hierarchical model for learning natural scene categories. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 2, pages 524–531. IEEE, 2005. 54

[35] David Feil-Seifer and Maja J Matarić. Defining socially assistive robotics. In Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on, pages 465–468. IEEE, 2005. 2, 3, 13

[36] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots: Concepts. Design, and Applications, 10:16–17, 2002. 2

[37] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots. Robotics and autonomous systems, 42(3):143–166, 2003. 2, 3

[38] Moshe Gabel, Ran Gilad-Bachrach, Erin Renshaw, and Assaf Schuster. Full body gait analysis with kinect. In Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, pages 1964–1967. IEEE, 2012. 46

[39] Hassan Ghasemzadeh, Vitali Loseu, and Roozbeh Jafari. Wearable coach for sport training: A quantitative model to evaluate wrist-rotation in golf. Journal of Ambient Intelligence and Smart Environments, 1(2):173–184, 2009. 64

[40] Barry R Greene, Alan O Donovan, Roman Romero-Ortuno, Lisa Cogan, Cliodhna Ni Scanaill, and Rose A Kenny. Quantitative falls risk assessment using the timed up and go test. Biomedical Engineering, IEEE Transactions on, 57(12):2918–2926, 2010. 45, 47

[41] T. Hamada, H. Okubo, K. Inoue, J. Maruyama, H. Onari, Y. Kagawa, and T. Hashimoto. Robot therapy as for recreation for elderly people with dementia -

game recreation using a pet-type robot -. In Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on, pages 174–179, Aug 2008. 10

[42] M. Heerink, B. Krose, V. Evers, and B. Wielinga. The influence of a robot's social abilities on acceptance by elderly users. In Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on, pages 521–526, Sept 2006. 16

[43] M. Heerink, B. Krose, V. Evers, and B. Wielinga. Observing conversational expressiveness of elderly users interacting with a robot and screen agent. In Rehabilitation Robotics, 2007. ICORR 2007. IEEE 10th International Conference on, pages 751–756, June 2007. 18

[44] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. Influence of social presence on acceptance of an assistive social robot and screen agent by elderly users. Advanced Robotics, 23(14):1909–1923, 2009. 18

[45] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. Assessing acceptance of assistive social agent technology by older adults: the almere model. International journal of social robotics, 2(4):361–375, 2010. xi, 102, 105, 107

[46] B Isaacs. Clinical and laboratory studies of falls in old people. prospects for prevention. Clinics in geriatric medicine, 1(3):513–524, 1985. 47

[47] C. Jayawardena, I. Kuo, C. Datta, R.Q. Stafford, E. Broadbent, and B.A. MacDonald. Design, implementation and field tests of a socially assistive robot for the elderly: Healthbot version 2. In Biomedical Robotics and Biomechatronics (BioRob), 2012 4th IEEE RAS EMBS International Conference on, pages 1837–1842, June 2012. 20

[48] Younbo Jung, Koay Jing Li, Ng Sihui Janissa, Wong Li Chieh Gladys, and Kwan Min Lee. Games for a better life: effects of playing wii games on the well-being of seniors in a long-term care facility. In Proceedings of the Sixth Australasian Conference on Interactive Entertainment, page 5. ACM, 2009. 4, 7

[49] Reza Kachouie, Sima Sedighadeli, Rajiv Khosla, and Mei-Tai Chu. Socially assistive robots in elderly care: a mixed-method systematic literature review. International Journal of Human-Computer Interaction, 30(5):369–393, 2014. 2

[50] Takayuki Kanda, Rumi Sato, Naoki Saiwaki, and Hiroshi Ishiguro. A two-month field trial in an elementary school for long-term human–robot interaction. Robotics, IEEE Transactions on, 23(5):962–971, 2007. 83

[51] Yukitaka Kawaguchi, Takanori Shibata, and Kazuyoshi Wada. The effects of robot therapy in the elderly facilities. Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 4(6):S133, 2010. 8

[52] Ronald C Kessler, Olga Demler, Richard G Frank, Mark Olfson, Harold Alan Pincus, Ellen E Walters, Philip Wang, Kenneth B Wells, and Alan M Zaslavsky. Prevalence and treatment of mental disorders, 1990 to 2003. New England Journal of Medicine, 352(24):2515–2523, 2005. 1

[53] Rajiv Khosla, Mei-Tai Chu, Reza Kachouie, Keiji Yamada, and Tomoharu Yamaguchi. Embodying care in matilda: an affective communication robot for the elderly in australia. In Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, pages 295–304. ACM, 2012. 14

[54] Rajiv Khosla, Mei-Tai Chu, Reza Kachouie, Keiji Yamada, Fujita Yoshihiro, and Tomoharu Yamaguchi. Interactive multimodal social robot for improving quality of care

of elderly in australian nursing homes. In Proceedings of the 20th ACM international conference on Multimedia, pages 1173–1176. ACM, 2012. 13

[55] Cory D Kidd and Cynthia Breazeal. Robots at home: Understanding long-term human-robot interaction. In Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on, pages 3230–3235. IEEE, 2008. 4, 7

[56] Quoc V Le, Will Y Zou, Serena Y Yeung, and Andrew Y Ng. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pages 3361–3368. IEEE, 2011. 73

[57] Emerson L Lesher and Jeffrey S Berryhill. Validation of the geriatric depression scale-short form among inpatients. Journal of clinical psychology, 1994. 93

[58] A Leu, D Ristic-Durrant, and A Graser. A robust markerless vision-based human gait analysis system. In Applied Computational Intelligence and Informatics (SACI), 2011 6th IEEE International Symposium on, pages 415–420. IEEE, 2011. 46

[59] Alexander V Libin and Elena V Libin. Person-robot interactions from the robopsychologists' point of view: the robotic psychology and robotherapy approach. Proceedings of the IEEE, 92(11):1789–1803, 2004. 16

[60] Magnus Lindwall, Mikael Rennemark, Anders Halling, Johan Berglund, and Peter Hassmén. Depression and exercise in elderly men and women: findings from the swedish national study on aging and care. Journal of aging and physical activity, 15(1):41, 2007. 2

[61] Christopher D Lorish and Richard Maisiak. The face scale: a brief, nonverbal method for assessing patient mood. Arthritis & Rheumatism, 29(7):906–909, 1986. 43

[62] S Mathias, US Nayak, Bernard Isaacs, et al. Balance in elderly patients: the" get-up and go" test. Archives of physical medicine and rehabilitation, 67(6):387–389, 1986. 47

[63] Yosuke Matsusaka, Hiroyuki Fujii, Token Okano, and Isao Hara. Health exercise demonstration robot taizo and effects of using voice command in robot-human collaborative demonstration. In Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on, pages 472–477. IEEE, 2009. 3, 7

[64] Oli Mival, Steward Cringean, and David Benyon. Personification technologies: Developing artificial companions for older people. CHI Fringe, Austria, 2004. 11

[65] Sarah Shizuko Morimoto, Bruce E Wexler, Jiacheng Liu, Willie Hu, Joanna Seirup, and George S Alexopoulos. Neuroplasticity-based computerized cognitive remediation for treatment-resistant geriatric depression. Nature communications, 5, 2014. 41

[66] T. Nakashima, G. Fukutome, and N. Ishii. Healing effects of pet robots at an elderly-care facility. In Computer and Information Science (ICIS), 2010 IEEE/ACIS 9th International Conference on, pages 407–412, Aug 2010. 15

[67] Ellinor Nordin, Erik Rosendahl, and Lillemor Lundin-Olsson. Timed up & go test: reliability in older people dependent in activities of daily living?focus on cognitive state. Physical therapy, 86(5):646–655, 2006. 47

[68] Hamed Pirsiavash, Carl Vondrick, and Antonio Torralba. Assessing the quality of actions. In Computer Vision–ECCV 2014, pages 556–571. Springer, 2014. 58, 73, 74, 76, 77, 78

[69] Diane Podsiadlo and Sandra Richardson. The timed" up & go": a test of basic functional mobility for frail elderly persons. Journal of the American geriatrics Society, 39(2):142–148, 1991. 45, 47

[70] Ronald Poppe. Vision-based human motion analysis: An overview. Computer vision and image understanding, 108(1):4–18, 2007. 45

[71] Alberto GRECO Giuseppe ANERDI Guido RODRIGUEZ. Acceptance of year animaloid robot as a starting points for cognitive stimulators supporting elders with cognitive impairments. 11

[72] Selma Sabanovic, Casey C Bennett, Wan-Ling Chang, and Laszlo Huber. Paro robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. In Rehabilitation Robotics (ICORR), 2013 IEEE International Conference on, pages 1–6. IEEE, 2013. 8

[73] Kamya Sakairi. Research of robot-assisted activity for the elderly with senile dementia in a group home. In SICE 2004 Annual Conference, volume 3, pages 2092–2094. IEEE, 2004. 11

[74] Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. Acoustics, Speech and Signal Processing, IEEE Transactions on, 26(1):43–49, 1978. 59

[75] Stan Salvador and Philip Chan. Fastdtw: Toward accurate dynamic time warping in linear time and space. In KDD workshop on mining temporal and sequential data. Citeseer, 2004. ix, 67

[76] J Reynaldo A Santos. Cronbachs alpha: A tool for assessing the reliability of scales. Journal of extension, 37(2):1–5, 1999. 102

[77] Mohammad Shayganfar, Charles Rich, and Candace L Sidner. A design methodology for expressing emotion on robot faces. In Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on, pages 4577–4583. IEEE, 2012. 32

[78] RQ Stafford, Elizabeth Broadbent, Chandimal Jayawardena, Ulrike Unger, I Han Kuo, Aleksandar Igic, Richie Wong, Ngaire Kerse, C Watson, and Bruce A MacDonald. Improved robot attitudes and emotions at a retirement home after meeting a robot. In RO-MAN, 2010 IEEE, pages 82–87. IEEE, 2010. 20

[79] Erik E Stone and Marjorie Skubic. Evaluation of an inexpensive depth camera for passive in-home fall risk assessment. In PervasiveHealth, pages 71–77, 2011. 46

[80] Toshiyo Tamura, Satomi Yonemitsu, Akiko Itoh, Daisuke Oikawa, Akiko Kawakami, Yuji Higashi, Toshiro Fujimooto, and Kazuki Nakajima. Is an entertainment robot useful in the care of elderly people with severe dementia? The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 59(1):M83–M85, 2004. 11

[81] Adriana Tapus. Improving the quality of life of people with dementia through the use of socially assistive robots. In Advanced Technologies for Enhanced Quality of Life, 2009. AT-EQUAL'09., pages 81–86. IEEE, 2009. 12, 13

[82] Adriana Tapus, Juan Fasola, and Maja J Matarić. Cognitive assistance and physical therapy for dementia patients using socially assistive robots. Social Interaction with Intelligent Indoor Robots (SI3R), 2008. 13

[83] Adriana Tapus, Cristian Tapus, and Maja Mataric. Music therapist robot: A solution for helping people with cognitive impairments. In IJCAI Workshop on Intelligent Systems for Assisted Cognition, July, Pasadena, CA, 2009. 13

[84] Adriana Tapus, Cristian Ţăpuş, and Maja Matarić. Music therapist robot for individuals with cognitive impairments. In Human-Robot Interaction (HRI), 2009 4th ACM/IEEE International Conference on, pages 297–298. IEEE, 2009. 13

[85] Adriana Tapus, Cristian Tapus, and Maja Mataric. Music therapist robot for people suffering from dementia: Longitudinal study. Alzheimer's & Dementia, 5(4):P338, 2009. 13

[86] Adriana Tapus, Cristian Tapus, and Maja Matarić. The role of physical embodiment of a therapist robot for individuals with cognitive impairments. In Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on, pages 103–107. IEEE, 2009. 12

[87] Yin-Leng Theng, Amirrudin Bin Dahlan, Meutia Latifah Akmal, and Thant Zin Myint. An exploratory study on senior citizens' perceptions of the nintendo wii: the case of singapore. In Proceedings of the 3rd International Convention on Rehabilitation Engineering & Assistive Technology, page 10. ACM, 2009. 4, 7

[88] Eduardo Velloso, Andreas Bulling, and Hans Gellersen. Motionma: motion modelling and analysis by demonstration. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 1309–1318. ACM, 2013. 61

[89] Kazuyoshi Wada, Yousuke Ikeda, Kaoru Inoue, and Reona Uehara. Development and preliminary evaluation of a caregiver's manual for robot therapy using the therapeutic seal robot paro. In RO-MAN, 2010 IEEE, pages 533–538. IEEE, 2010. 107

[90] Kazuyoshi Wada and Takanori Shibata. Living with seal robotsits sociopsychological and physiological influences on the elderly at a care house. Robotics, IEEE Transactions on, 23(5):972–980, 2007. 8

[91] Kazuyoshi Wada, Takanori Shibata, and Yukitaka Kawaguchi. Long-term robot therapy in a health service facility for the aged-a case study for 5 years. In Rehabilitation Robotics, 2009. ICORR 2009. IEEE International Conference on, pages 930–933. IEEE, 2009. 10

[92] Kazuyoshi Wada, Takanori Shibata, Tomoko Saito, Kayoko Sakamoto, and Kazuo Tanie. Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged. In Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on, pages 2785–2790. IEEE, 2005. 43

[93] Carolee J Winstein, J Philip Miller, Sarah Blanton, Edward Taub, Gitendra Uswatte, David Morris, Deborah Nichols, and Steven Wolf. Methods for a multisite randomized trial to investigate the effect of constraint-induced movement therapy in improving upper extremity function among adults recovering from a cerebrovascular stroke. Neurorehabilitation and Neural Repair, 17(3):137–152, 2003. 4

[94] Dennis Wollersheim, Monika Merkes, Nora Shields, Pranee Liamputtong, Lara Wallis, Fay Reynolds, Lee Koh, et al. Physical and psychosocial effects of wii video game use among older women. International Journal of Emerging Technologies and Society, 8(2):85–98, 2010. 4, 7

[95] Yi Yang and Deva Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pages 1385–1392. IEEE, 2011. 73, 74

[96] Xiao Zhang, Ali Mollahosseini, B Kargar, H Amir, Evan Boucher, Richard M Voyles, Rodney Nielsen, and MH Mahoor. ebear: An expressive bear-like robot. In Robot and

Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on, pages 969–974. IEEE, 2014. 21, 23