University of Denver

Digital Commons @ DU

Electronic Theses and Dissertations | Graduate Studies

2023

Artificial Emotional Intelligence in Socially Assistive Robots

Hojjat Abdollahi
University of Denver

Follow this and additional works at: https://digitalcommons.du.edu/etd

Part of the Artificial Intelligence and Robotics Commons, Electrical and Computer Engineering Commons, and the Robotics Commons

Recommended Citation

Abdollahi, Hojjat, "Artificial Emotional Intelligence in Socially Assistive Robots" (2023). *Electronic Theses and Dissertations*. 2217.
https://digitalcommons.du.edu/etd/2217

This Dissertation is brought to you for free and open access by the Graduate Studies at Digital Commons @ DU. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ DU. For more information, please contact jennifer.cox@du.edu,dig-commons@du.edu.

Artificial Emotional Intelligence in Socially Assistive Robots

Abstract

Artificial Emotional Intelligence (AEI) bridges the gap between humans and machines by demonstrating empathy and affection towards each other. This is achieved by evaluating the emotional state of human users, adapting the machine's behavior to them, and hence giving an appropriate response to those emotions. AEI is part of a larger field of studies called Affective Computing. Affective computing is the integration of artificial intelligence, psychology, robotics, biometrics, and many more fields of study. The main component in AEI and affective computing is emotion, and how we can utilize emotion to create a more natural and productive relationship between humans and machines.

An area in which AEI can be particularly beneficial is in building machines and robots for healthcare applications. Socially Assistive Robotics (SAR) is a subfield in robotics that aims at developing robots that can provide companionship to assist people with social interaction and companionship. For example, residents living in housing designed for older adults often feel lonely, isolated, and depressed; therefore, having social interaction and mental stimulation is critical to improve their well-being. Socially Assistive Robots are designed to address these needs by monitoring and improving the quality of life of patients with depression and dementia. Nevertheless, developing robots with AEI that understand users' emotions and can reply to them naturally and effectively is in early infancy, and much more research needs to be carried out in this field.

This dissertation presents the results of my work in developing a social robot, called Ryan, equipped with AEI for effective and engaging dialogue with older adults with depression and dementia. Over the course of this research there has been three versions of Ryan. Each new version of Ryan is created using the lessons learned after conducting the studies presented in this dissertation. First, two human-robot-interaction studies were conducted showing validity of using a rear-projected robot to convey emotion and intent. Then, the feasibility of using Ryan to interact with older adults is studied. This study investigated the possible improvement of the quality of life of older adults. Ryan the Companionbot used in this project is a rear-projected lifelike conversational robot. Ryan is equipped with many features such as games, music, video, reminders, and general conversation. Ryan engages users in cognitive games and reminiscence activities. A pilot study was conducted with six older adults with early-stage dementia and/or depression living in a senior living facility. Each individual had 24/7 access to a Ryan in his/her room for a period of 4-6 weeks. The observations of these individuals, interviews with them and their caregivers, and analysis of their interactions during this period revealed that they established rapport with the robot and greatly valued and enjoyed having a companionbot in their room.

A multi-modal emotion recognition algorithm was developed as well as a multi-modal emotion expression system. These algorithms were then integrated into Ryan. To engage the subjects in a more empathic interaction with Ryan, a corpus of dialogues on different topics were created by English major students. An emotion recognition algorithm was designed and implemented and then integrated into the dialogue management system to empathize with users based on their perceived emotion. This study investigates the effects of this emotionally intelligent robot on older adults in the early stage of depression and dementia. The results of this study suggest that Ryan equipped with AEI is more engaging, likable, and attractive to users than Ryan without AEI. The long-term effect of the last version of Ryan (Ryan V3.0) was studied in a study involving 17 subjects from 5 different senior care facilities. The participants in this study experienced a general improvement in their cognitive and depression scores.

Document Type

Dissertation

Degree Name
Ph.D.

Department
Computer Science and Engineering

First Advisor
Mohammad H. Mahoor

Second Advisor
Kimon Valavanis

Third Advisor
Timothy Sweeny

Keywords
Artificial intelligence, Emotional intelligence, Social robotics

Subject Categories
Artificial Intelligence and Robotics | Computer Engineering | Computer Sciences | Electrical and Computer Engineering | Engineering | Robotics

Publication Statement
Copyright is held by the author. User is responsible for all copyright compliance.

This dissertation is available at Digital Commons @ DU: https://digitalcommons.du.edu/etd/2217

ARTIFICIAL EMOTIONAL INTELLIGENCE IN SOCIALLY ASSISTIVE ROBOTS

A Dissertation

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

Doctor of Philosophy

by

Hojjat Abdollahi

June 2023

Advisor: Mohammad H. Mahoor, Ph.D.

Author: Hojjat Abdollahi
Title: ARTIFICIAL EMOTIONAL INTELLIGENCE IN SOCIALLY ASSISTIVE ROBOTS
Advisor: Mohammad H. Mahoor, Ph.D
Degree Date: June 2023

Abstract

Artificial Emotional Intelligence (AEI) bridges the gap between humans and machines by demonstrating empathy and affection towards each other. This is achieved by evaluating the emotional state of human users, adapting the machine's behavior to them, and hence giving an appropriate response to those emotions. AEI is part of a larger field of studies called *Affective Computing*. Affective computing is the integration of artificial intelligence, psychology, robotics, biometrics, and many more fields of study. The main component in AEI and affective computing is emotion, and how we can utilize emotion to create a more natural and productive relationship between humans and machines.

An area in which AEI can be particularly beneficial is in building machines and robots for healthcare applications. Socially Assistive Robotics (SAR) is a subfield in robotics that aims at developing robots that can provide companionship to assist people with social interaction and companionship. For example, residents living in housing designed for older adults often feel lonely, isolated, and depressed; therefore, having social interaction and mental stimulation is critical to improve their well-being. Socially Assistive Robots are designed to address these needs by monitoring and improving the quality of life of patients with depression and dementia. Nevertheless, developing robots with AEI that understand users' emotions and can reply to them naturally and effectively is in early infancy, and much more research needs to be carried out in this field.

This dissertation presents the results of my work in developing a social robot, called Ryan, equipped with AEI for effective and engaging dialogue with older adults with de-

pression and dementia. Over the course of this research there has been three versions of Ryan. Each new version of Ryan is created using the lessons learned after conducting the studies presented in this dissertation. First, two human-robot-interaction studies were conducted showing validity of using a rear-projected robot to convey emotion and intent. Then, the feasibility of using Ryan to interact with older adults is studied. This study investigated the possible improvement of the quality of life of older adults. Ryan the Companionbot used in this project is a rear-projected lifelike conversational robot. Ryan is equipped with many features such as games, music, video, reminders, and general conversation. Ryan engages users in cognitive games and reminiscence activities. A pilot study was conducted with six older adults with early-stage dementia and/or depression living in a senior living facility. Each individual had 24/7 access to a Ryan in his/her room for a period of 4-6 weeks. The observations of these individuals, interviews with them and their caregivers, and analysis of their interactions during this period revealed that they established rapport with the robot and greatly valued and enjoyed having a companionbot in their room.

A multi-modal emotion recognition algorithm was developed as well as a multi-modal emotion expression system. These algorithms were then integrated into Ryan. To engage the subjects in a more empathic interaction with Ryan, a corpus of dialogues on different topics were created by English major students. An emotion recognition algorithm was designed and implemented and then integrated into the dialogue management system to empathize with users based on their perceived emotion. This study investigates the effects of this emotionally intelligent robot on older adults in the early stage of depression and dementia. The results of this study suggest that Ryan equipped with AEI is more engaging, likable, and attractive to users than Ryan without AEI. The long-term effect of the last version of Ryan (Ryan V3.0) was studied in a study involving 17 subjects from 5 different senior care facilities. The participants in this study experienced a general improvement in their cognitive and depression scores.

Acknowledgements

I would first and foremost like to express my deepest gratitude to my caring and supportive advisor, Dr. Mohammad H. Mahoor, for his direction, expertise, and overwhelming generosity. Dr. Mahoor, you have unquestionably shaped my research, my career, and my life in the most positive ways. Thank you!

Next, I extend my sincere thanks to my esteemed committee members, Dr. Kimon Valavanis, Dr. Matthew Rutherford, and Dr. Timothy Sweeny. Your valuable time, constructive feedback, and support have helped me bring this project to the finish line.

There are many individuals who I would like to acknowledge for their support throughout these years. To my colleagues, Ali Mollahosseini, Ali Pourramezan Fard, Joshua Lane, Rohola Zandie, senior care facilities and their hardworking staff, specifically Eaton Senior Communities and Sarah Schoeder, and sweet and kind senior care residents who graciously agreed to participate in my studies. Thank you all!

And finally, I would like to thank my partner, Elene. I cannot fully express how grateful I am for your love, support, and encouragement. I would also like to express my gratitude to my family, who have done everything they can to further my education.

Funding disclaimer: Development of Ryan the CompanionBot was supported through a SBIR grant awarded to DreamFace Technologies (DFT), LLC by the National Science Foundation (NSF) under award IIP-1548956. Parts of the research reported in this manuscript were supported by the National Institute on Aging (NIA) of the National Institutes of Health (NIH) under SBIR award R44AG059483 awarded to DFT. The study in Chapter 5 was supported by the NIH/NIA under SBIR award R44AG066439 awarded to DFT. The content of this manuscript is solely the responsibility of me (the author) and does not necessarily represent the official views of the NIH or NSF.

Table of Contents

List of Tables

List of Figures

Acronyms

AD Alzheimer's Disease.

ADRD Alzheimer's Disease and Related Dementia.

AEI Artificial Emotional Intelligence.

AI Artificial Intelligence.

AIML Artificial Intelligence Markup Language.

CNN Convolutional Neural Network.

CR Co-Presence Robot.

DFT DreamFace Technologies. LLC..

DNN Deep Neural Network.

EI Emotional Intelligence.

FER Facial Expression Recognition.

GDS Geriatric Depression Scale.

GT Ground Truth.

HRI Human Robot Interaction.

iCBT Internet-based Cognitive Behavioural Therapy.

IRB Institutional Review Board.

LMM Linear Mixed-effects Model.

NIA National Institute on Aging.

NIH National Institute of Health.

NSF National Science Foundation.

PHQ-9 Patient Health Questionnaire-9.

SA Sentiment Analysis.

SAR Socially Assistive Robotics.

SBIR Small Business Innovation Research.

SLUMS Saint Louis University Mental Status.

TR Tele-Presence Robot.

UI User Interface.

VA Virtual Agent.

WOZ Wizard-Of-Oz.

Chapter 1

Introduction

The research described in this dissertation focuses on the potential benefits of artificially emotionally intelligent social robots in elder care, specifically for individuals with early-stage Alzheimer's Disease (AD)/Alzheimer's Disease and Related Disorders (ADRD) and mild depression. The goal is to explore how these robots can improve the social, mental, and physical well-being of older adults and potentially slow down the progression of these debilitating diseases. To achieve this, we develop a social robot, equip it with tools for a more natural and affect-aware interaction, and study its effect on older adults living in senior care facilities in the Denver Metro Area.

This introductory chapter provides a concise overview of the motivation behind this study, highlighting the implications of demographic changes and the aging population, as well as the potential role of robots and artificial emotional intelligence in addressing the challenges associated with elder care.

Finally, we then describe the focus of our work and provide a brief overview of the structure and organization of this dissertation.

1.1 Motivation

1.1.1 AD/ADRD

More than 6 million Americans are living with Alzheimer's disease (AD), and it is expected that the number will rise to 13.8 million by 2050 as the population ages (Hebert et al., 2013). AD/ADRD is the sixth leading cause of death among adults 65 years of age and older, with deaths more than doubling between 2000 and 2019 (Alz; Hebert et al., 2013). Figure 1.1, shows the facts and figures for 2022 by Alzheimer's association. Common symptoms of AD and AD related dementia (ADRD) include cognitive decline, short-term memory loss, changes in mood, depression, communication difficulties, loss of interest in hobbies or activities, and repetitive behavior. In 2023, Alzheimer's and other dementias will cost US $345 billion with projected annual costs of over $1 trillion by 2050 (Alz). Unfortunately, to date, there are no effective treatments available to cure AD and dementia. Traditional disease management methods have shown limited success in treating AD or mitigating its symptoms; The drug failure rate for AD is currently 99.6% (compared to 81% for cancer) (Cummings, 2018). Although there is no cure for dementia, the Alzheimer's Association emphasizes the importance of helping individuals with AD keep their brain active through social interaction, music therapy, reminiscence therapy, and other cognitive activities, in addition to pharmacological treatment (Alz).

Due to the growing demand for the care and treatment of elderly people with dementia, healthcare personnel and caregivers are physically and emotionally taxed and actively seek new methods to assist the growing number of people. According to Plunkett, "Healthcare is one of the largest and fastest growing industries in the world, and virtually all government and private health initiatives that pay for health care are desperately seeking ways to improve patient care outcomes". By 2030, the global demand for health workers will rise to 80 million workers, while the supply is expected to reach 65 million, resulting in a worldwide shortage of 15 million health workers (Liu et al., 2017).

1.1.2 Depression

Depression is the most common type of mental disorder in the United States. Depression increases the risk of many physical health problems, particularly long-lasting conditions such as diabetes, heart disease, and stroke. It causes severe symptoms that affect how you feel, think, and handle daily activities, such as sleeping, eating, or working. Research suggests that genetic, biological, environmental and psychological factors play a role in depression (Saveanu and Nemeroff, 2012). According to the National Institute of Mental Health, in 2020: 1) an estimated 21.0 million adults in the United States had at least one major depressive episode. This number represented 8.4% of all US adults. 2) The prevalence of major depressive episodes was higher among adult women (10.5%) compared to men (6.2%) (NIMH, 2020).

Depression can cause tremendous challenges and burdens for individuals and families. According to Greenberg et al. (2021) the economic burden of major depressive disorder among US adults was estimated at $236 billion in 2018, an increase of more than 35% since 2010 (values for 2020). Depression can occur at any age, but often begins in adulthood. Depression is now recognized to occur in children and adolescents, although it sometimes presents with more prominent irritability than low mood. Many chronic mood and anxiety disorders in adults start as high levels of anxiety in children. Depression, especially in midlife or older adults, can co-occur with other serious medical illnesses, such as diabetes, cancer, heart disease, Alzheimer's disease, and Parkinson's disease. These conditions are often worse when depression is present, and research suggests that people with depression and other medical conditions tend to have more severe symptoms of both illnesses. Thoughts of death or suicide, or suicide attempts, are common symptoms of depression.

1.2 Socially Assistive Robots

As the global population ages, the demand for elder care increases and the need for innovative solutions becomes more pressing. Social robots have emerged as a promising technology to address some of the challenges associated with elder care, such as social isolation, depression, and cognitive decline.

Socially intelligent robotics is a rapidly emerging field aiming to design robots that are able to communicate and interact with humans in a socially acceptable way (Breazeal, 2005; Dautenhahn, 2007). They often achieve positive outcomes in diverse applications such as education, health-care, quality of life, entertainment, communication, and tasks requiring collaborative teamwork (Breazeal et al., 2016). These robots are becoming an integrated part of our daily lives. For social robots to be able to communicate with us naturally, they need to be more affect-aware.

Affective computing is the integration of artificial intelligence, psychology, robotics, biometrics, and many other fields of study. Affective computing allows us to interact with machines and robots using our emotions (Yonck, 2020). It may be difficult to formally define an emotion, but it is evident that emotions are at the center of human experience. A famous question posed by Minsky (1988) asks: “the question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without any emotions.” Affective computing is a very natural progression in our ongoing efforts to build technologies that operate increasingly on human terms, rather than the other way around. Damasio (1994) presented some neurological evidence to prove that emotions do in fact play an active and important role in the human decision-making process. The interaction between the emotional process and the cognitive process may explain why humans excel in making decisions based on incomplete information, “acting on our gut feelings”. This in turn was the reason for emergence of the term “emotional intelligence”.

With the recent growth in the adoption of Artificial Intelligence (AI) technology in a variety of applications and disciplines, every aspect of our lives will soon be affected by AI. Personal assistants in our pockets, robots in our homes and workplaces, as well as self-driving cars on the streets, are just a few examples of the ubiquity of AI. The AI community has mostly focused on making smarter and more intelligent systems that are capable of solving hard technical problems, though the need for emotionally intelligent systems that understand users’ feelings and can connect with them in a natural and welcoming manner is growing rapidly.

An area in which AEI can be particularly beneficial is in building machines and robots for healthcare applications. Socially Assistive Robotics (SAR) is a niche field in robotics that aims at developing robots that can provide companionship to assist people with social interaction and companionship. For instance, residents living in housing designed for older adults often feel lonely, isolated, depressed and hence having social interaction and mental stimulation is critical for improving their well-being. Socially Assistive Robots, such as Ryan and Pepper are designed to address these needs by monitoring and improving the quality of life of patients with depression and dementia. Nevertheless, developing robots with AEI that understand users’ emotions and can reply to them naturally and effectively is in early infancy, and much more research needs to be carried out in this field.

A futuristic version of an emotionally intelligent machine was depicted in the film Her (2013). After only 10 years of development in AI, especially affective computing, an emotional relationship with a machine does not seem far-fetched anymore. While humanoid robots such as Sophia are capable of showing emotions using a prosthetic face, a truly emotionally intelligent robot needs to be able to perceive the user's emotion and mental state, devise an emotionally appropriate response, and then convey it while expressively.

1.3 The outline

This dissertation provides a research investigation into the potential use of socially assistive robots in elder care, with a specific emphasis on the design and development of an emotionally intelligent robot named Ryan. Chapter 2 covers the related work and literature review. It studies the robots used in elder care and other studies in artificial emotional intelligence. Chapter 3 introduces Ryan, the Companionbot, that serves as the focus of this research. Three versions of Ryan are explained in detail in this chapter. Chapter 4 presents the preliminary human-robot-interaction studies conducted to explore various features of Ryan in a lab environment. Chapter 5 delves into the use of Ryan in elder care and examines the feasibility of a robotic companion and the potential for an emotional bond between humans and a robot. Finally, Chapter 6 presents the design and development of a multi-modal emotion recognition algorithm and a multi-modal emotion expression system that are integrated into Ryan. Furthermore, this chapter presents the outcomes of two separate studies that were conducted using the emotionally intelligent version of Ryan. This chapter offers an in-depth analysis of the findings from these studies, which serve to further demonstrate the efficacy and potential of incorporating emotional intelligence into the design of socially assistive robots.

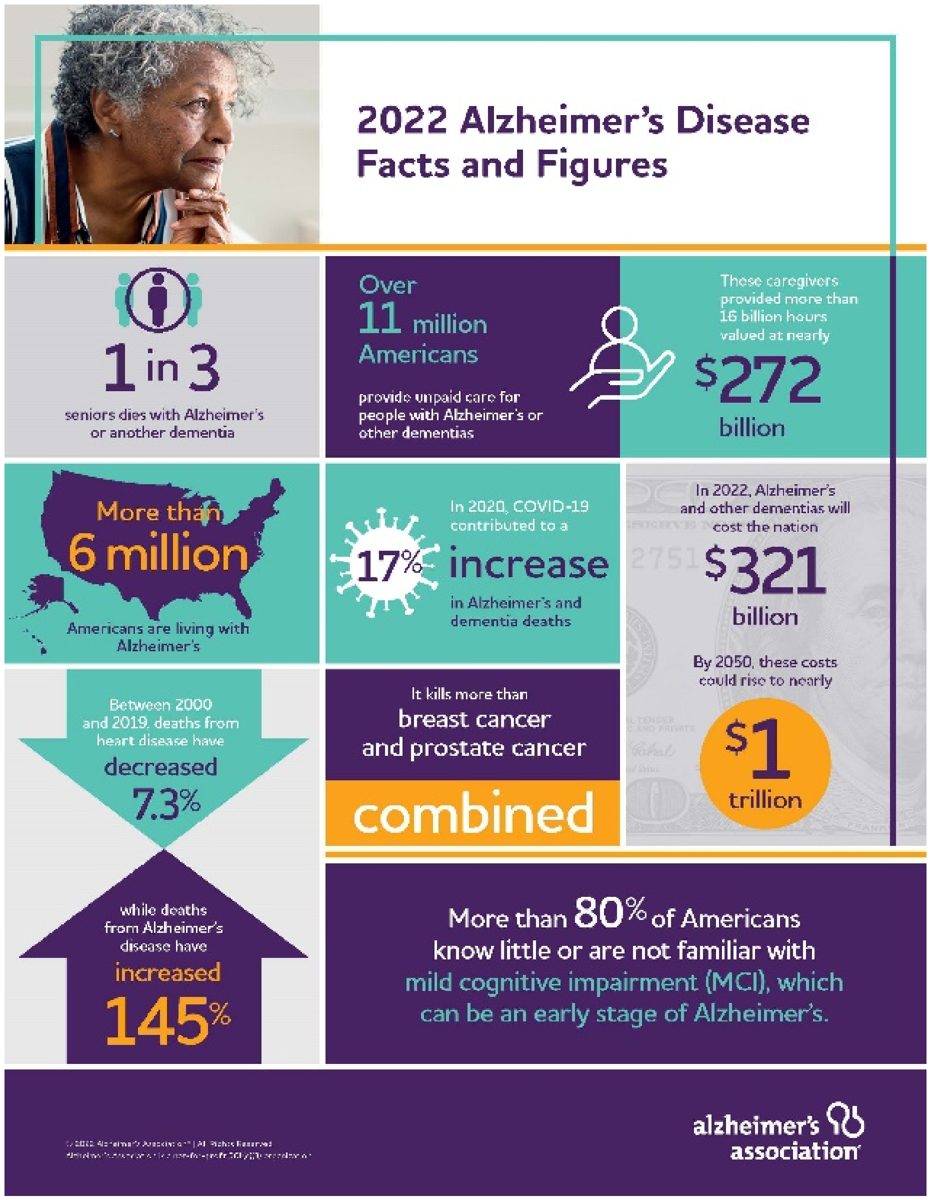


Figure 1.1: 2022 Alzheimer's disease facts and figures.

Chapter 2

Literature Review and Related Works

In recent years, there has been an increase in the use of robots not only in industrial fields, but also in other areas such as schools (Conti et al., 2018), homes (Cavallo et al., 2014), hospitals (D'Onofrio et al., 2018), rehabilitation centers (Loi et al., 2018), and in senior care facilities (Khosla et al., 2012).

There are three types of robots used in healthcare: **1)**: Physically assistive robots, such as sophisticated wheelchairs and surgical robots, **2)**: Socially interactive robots, such as spoken-dialog-enabled receptionists, and **3)**: Socially assistive robots, such as Pepper (Figure 2.1.i), that are designed to monitor and improve the quality of life of users. Our focus in this dissertation is on the last type.

In this dissertation, I have focused on social robots used in healthcare, specifically with older adults. The use of socially assistive robots (SAR) to help older adults has recently become more relevant due to the increase in the number of elderly people, the decrease in the cost of technology, and recent advances in artificial intelligence Leite (2015). Nursing home residents live alone with disabilities, while in most cases their cognitive abilities are degraded due to old age or various types of dementia Kotwal et al. (2016). Studies suggest that social support for elderly people could improve their cognitive function (Zamora-

Macorra et al., 2017). Using SARs with a focus on the socialization aspect of Human-Robot Interaction (HRI) is a viable option to reduce the burden on caregivers while providing companionship to elderly people, improving their quality of life and avoiding depression and further degradation of their mental abilities.

Figure 2.1: Robots used in social robotics studies.

Wada et al. (2003) used the Paro (Figure 2.1.f) robot to study the long-term effect of social robots on residents of a senior care center. The results indicated that the residents established a relationship with the robot, developed stronger social bonds among themselves, and also maintained a lower stress level. However, Paro lacks the ability to talk and communicate. It is shown that to be accepted more easily, a social robot should be communicative (Heerink et al., 2006) and must employ a form of communication with which humans are habituated (Krämer et al., 2012). Paro does not have any human-like features.

Studies show that humans feel closer to a social robot when they interact with it in a one-on-one setting (Lee et al., 2005). It has also been shown that to build a relationship with robots, humans use principles that are more in line with the human-human interaction than the human-robot interaction (Krämer et al., 2012). These studies suggest that a robot can substitute human companionship or at least have a higher chance of making a bond if the robot looks and acts like a human. Anthropomorphic characteristics in social robots can facilitate human social understanding (Breazeal, 2000; Duffy et al., 2002) and are important in the development of a meaningful social interaction between robots and people (Duffy, 2003)

Another key aspect of having a robot as a companion is continuous (uninterrupted) companionship, which means having access to the robot at all times. The majority of studies on human-robot interaction (HRI) are often brief and conducted in public spaces, or focus on specific domains such as education (Michaelis and Mutlu, 2017), healthcare (Bodala et al., 2021; Robinson et al., 2020; Van Maris et al., 2020), or rehabilitation (Céspedes et al., 2021a,b). However, conducting longitudinal studies that involve physically embodied social robots in users' homes to examine repeated interactions remains rare due to the logistical and cost-related challenges associated with situating these devices in domestic settings.

To gain a complete understanding of how humans adapt to social robots and how their perceptions and behavior evolve over time, it is crucial to conduct longitudinal studies (Leite et al., 2013a). Autonomy plays a crucial role in achieving uninterrupted companionship that enables a longitudinal study. Most studies conducted with social robots in elder care are performed in a Wizard-Of-Oz (WOZ) manner (Vardoulakis et al., 2012), or were limited to a specific scenario (Pineau et al., 2003). Vardoulakis et al. (2012) designed an experiment to study the long-term social companion of older adults. They used a WOZ method and the subject had a robot at home for one week. However, since the robot was remotely controlled by an operator, the subject interacted with the robot for only one hour every day. Employing the WOZ method forces the subjects to use the robot at a specific time of the day for a short period, which resembles visiting a friend rather than having a companion at home. Social robots such as Paro are autonomous and provide continuous companionship, but lack the ability of having a robust social interaction such as spoken dialog and an expressive face.

Studies suggest that social robots are promising tools for delivering and improving mental health interventions (Robinson et al., 2019), supporting rehabilitation (Feingold Polak and Tzedek, 2020), and providing physical and social support (Henschel et al., 2021) in various settings. Research has shown that social robots can help minimize social tensions and anxieties(Nomura et al., 2020), particularly for those with social anxiety, and can serve as interventions for social anxiety(Rasouli et al., 2022).

Furthermore, the COVID-19 pandemic has highlighted the potential of social robots as assistive technology, as they can perform tasks such as taking temperature, food and supply delivery, providing companionship, and mediating social interactions (Henschel and Cross, 2020; Scassellati and Vázquez, 2020; Yang et al., 2020).

Deep social interaction is required when dealing with older adults with dementia. Different robots such as Paro, Nao, and Zeno (Figure 2.1.a,e,f) have been used in studies on the care of elderly people with dementia (Mordoch et al., 2013). Most of the robots used in these studies have not been built with the social aspect in mind. But to be able to communicate with older adults with AD/ADRD and try to engage them in conversations and games, we need a robot that has been designed to accomplish these social goals. Recently, several studies have investigated the incorporation of empathy into social robots (Alves-Oliveira et al., 2019; Leite et al., 2013b; Mollahosseini et al., 2018a; Paiva et al., 2005). In chapter 3, I introduce Ryan, a robot designed to be social, empathic, and emotionally intelligent.

Chapter 3

Ryan CompanioBot

Despite the tremendous efforts by many researchers in academia and industry to design and build realistic robotic heads, current robots have yet to reach the perceptive and emotional verbal and nonverbal social capabilities of humans. These social capabilities, which include the ability to engage users in natural spoken dialog, interpreting users' affect states, and respond effectively to them through speech and facial expressions, are necessary for rich and robust interaction with human beings. Social robots such as Paro (Kidd et al., 2006) have the robustness and cost effectiveness for large scale, unattended user trials, but lack the sophistication for deep social interaction. Social robots such as Simon (Simon) possess state-of-the-art capabilities for social interaction, but are too expensive and maintenance-intensive.

The robot used in this study is Ryan Companionbot (Ryan) which is based on Expressionbot (Mollahosseini et al., 2014a). Ryan has been developed in DreamFace Technologies, LLC. with the social aspect of HRI in mind. This robot has an emotive and expressive face with accurate visual speech. Ryan can maintain a spoken dialog, recognize expressions on the user's face, and it is equipped with a screen on its torso with features such as cognitive games, music player, narrated photo album, and video player.

To keep the subject engaged for a long period of time, the SARs must be personalized (Castellano et al., 2008a). Thus, Ryan can be customized for each user. To increase intimacy and invoke rapport, users can choose a name for the robot. It is worth mentioning that in the study in Chapter 5 one subject named the robot after his late wife. In that study we left the Ryan with the subject for 4 weeks. Leaving the robot in an older adult's home and having 24/7 access to the robot may cause them to lose motivation. To provoke subjects to act on intrinsic motivation, we had to define tasks and modify Ryan to be enjoyable and not repetitive.

After a while that the user exhausts all of the features of the robot, they will lose interest in interacting with the robot. It is shown that the novelty effect of SARs disappears quickly (You et al., 2006). As the novelty aspect wears off, the social effect could also decrease (Fernaeus et al., 2010). By endowing Ryan with a character and a sense of humor on top of various other features implemented into Ryan, we keep the subjects interested to interact with the robot for a long period.

In the next sections, we will explain the hardware and software of different versions of Ryan.

3.1 Ryan V1.0

3.1.1 Hardware

Ryan V1.0 hardware is designed with three main components (Figure 3.1): 1) the head projection system, 2) the neck mechanism, and 3) the torso.

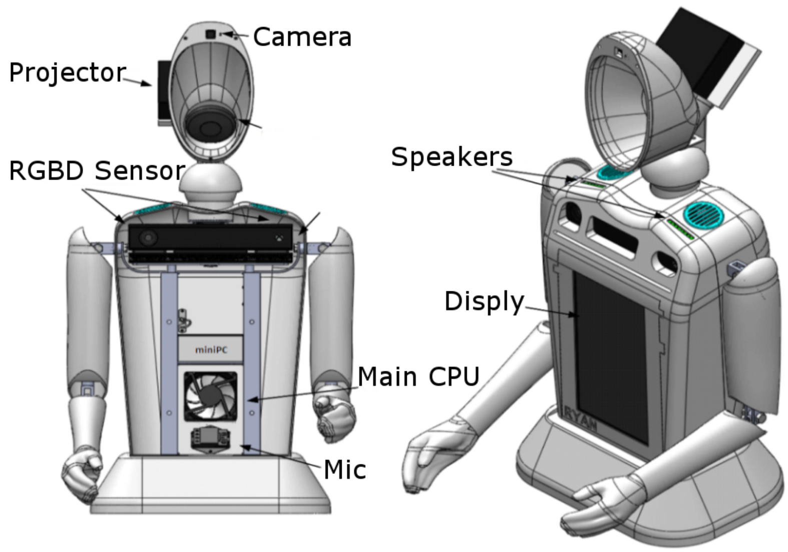


Figure 3.1: Ryan hardware.

3.1.2 Head Projection System

Using a large number of actuators to build a human-like robotic face capable of showing different emotions and visual speech is difficult and expensive (Mollahosseini et al., 2014b). To avoid the tremendous effort required to develop a robotic head capable of having accurate visual speech, state-of-the-art character animation technology was used to produce an avatar. Using rear projection optics, the head projection system displays the animated avatar on a mask. This system also allows us to further customize the appearance of the robot. Consult the work of Mollahosseini *et al*. (Mollahosseini et al., 2014b) for more details on the projection system.

Neck Mechanism

The movement of the head to track faces and head gestures is controlled by the neck mechanism, a two-degree-of-freedom pan/tilt unit. Having only two degrees of freedom keeps the system simple and suffices for face tracking. The neck has a range of motion of

30°of flexion and extension (±30° pitch) and 180°lateral rotation (±90°yaw). This range allows the head to track the user anywhere in front of the robot.

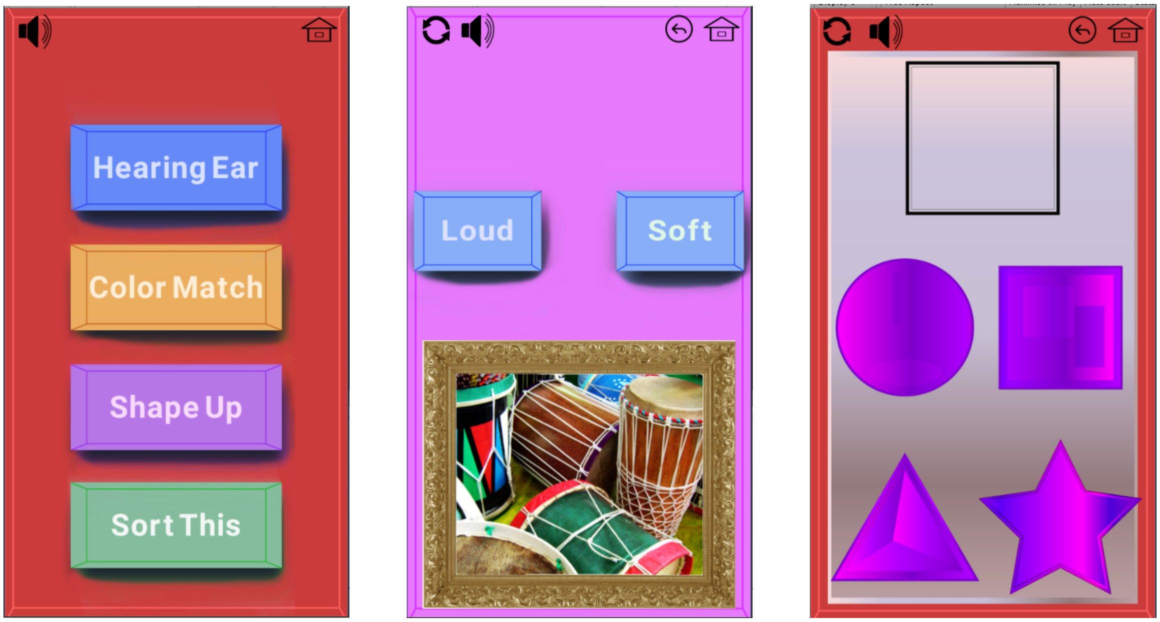
Torso

The main computer, an RGBD camera, a touch screen display, and power supplies are enclosed inside the torso. Adding a touch screen to the robot added a new way of interacting with Ryan (touch) and also it added the feature to be able to display more information to the user. The display was used for cognitive games, music player, video player, and the narrated photo album. The RGBD camera enables us to have a 3D view of the environment for better tracking the user and also for future studies on activity recognition.

3.1.3 Software

To make Ryan an intelligent and sociable robot that can understand human language and communicate through spoken dialog, a series of features have been implemented in the robot. Ryan must be able to find the user in the environment, read the user's facial expression, understand the user's speech, generate an appropriate response, and say it to the user through audio, accompanied with visual speech while showing a relevant expression on the face. Ryan is also able to communicate with users through the touch screen in the torso.

The Microsoft Kinect sensor V2.0 (Kinect) acts as the eyes of the system to constantly monitor user's activities and its face detection feature enables Ryan to find the subject in the room. For facial emotion recognition, Ryan uses the Intel RealSense SDK (Realsense) which provides seven basic facial expressions. Intel RealSense SDK is also used as the speech to text engine. Ryan uses the speech emotion recognition Aylien (Aylien) system which is an online natural language processing service for sentiment analysis of the user's



(a) Main Menu (b) Hearing Ear (c) Shape Up

Figure 3.2: Cognitive Games.

speech. A retrieval-based open dialog management systems available on the web (ChatBot/ Pandorabots (Pandorabots)) is used as the dialog manager.

To reduce subjects' cognitive abilities deterioration, we equipped Ryan with cognitive games focused on patients with dementia. Drugs are not the only method to treat mental diseases such as dementia, Alzheimer's disease, and depression. There exist alternative therapeutic methods such as talking therapies, life story and reminiscence work, and cognitive stimulation therapy for these diseases (Lawrence et al., 2012).

We designed four games (Figure 3.2). These games are based on Montessori-based activities (Judge et al., 2000) to help people suffering from dementia combat the disease. These visual games are simple and interactive with different levels of complexity. The game instructions were given by Ryan and the users could answer the questions either via voice commands or by pushing the buttons on the screen.

There is evidence that life story, photo albums, and reminiscence work, particularly when done one-on-one, can improve mood, well-being and some mental abilities such as memory (Lawrence et al., 2012). For each subject we collected about 15-20 old photos and the stories about the event in the photos, either from the participant or their close relatives. The photos are shown on the torso screen one by one, and the robot reads the story back to the user. Sometimes simple questions are asked to engage the user in the conversation.

Reminiscence and memory work also involves talking about things from the past, using prompts such as photos, familiar objects, or playing music. A video player application was created to randomly select and play videos from a list of YouTube video clips. The list contained URLs of short (4-5 minutes) YouTube videos queried based on the users' topics of interests (e.g. healthy foods, sports, and nature).

3.2 Ryan V2.0

Besides a full system overhaul of Ryan's aesthetics for a sleeker and a more visually appealing robot, we introduced three fundamental enhancements to Ryan's form and motion capabilities to improve social interaction and support between Ryan and the user. These enhancements focus on Ryan's arms, neck motion, and projection system. Ryan V2.0 was not used in any studies in this dissertation.

3.2.1 Hardware

Arm Enhancements: Although Ryan's purpose is to be socially and not physically assistive, we decided to include active arms in the second version. Active arms are used to engage users in physical exercise, which has been shown to elevate mood and have cognitive health benefits (Tseng et al., 2011). Both have a great impact on the quality of life of seniors with dementia. With active arms, Ryan is able to coach through demonstration of various exercises for the user and engage the user to perform the exercise in synchrony; adapting to the user's pace and providing positive reinforcement and feedback.

Safety feature: Although the incorporation of active arms creates opportunities for greater functionality, it also presents new challenges and risks; specifically, the risk of accidental collision of the arms with the user or their property. To mitigate these risks and ensure the safety of the users, a series of proximity sensors are installed on Ryan to ensure that Ryan only moves its arms if it is safe.

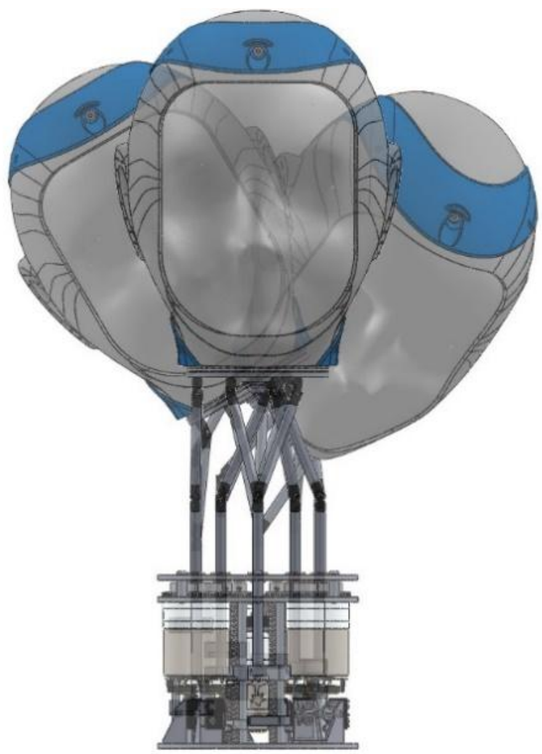


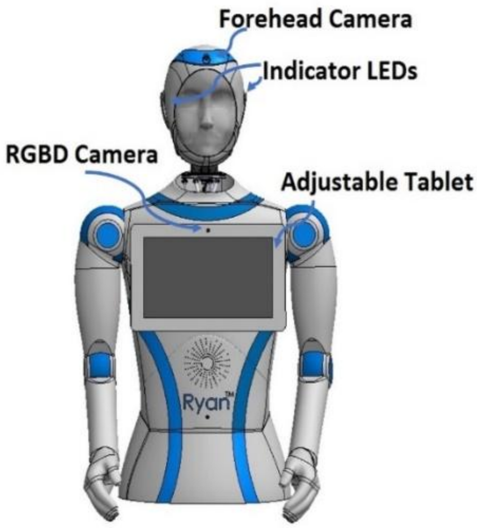
Figure 3.3: Six degree of freedom parallel neck mechanism providing natural head motion for more natural and emotive dialog.

Neck Enhancements: Studies have shown that head movement plays an important role in conversation. Head movements can signal familiarity and a sense of rapport with the conversational counterpart through gestures that communicate affirmation, impatience, disbelief, desire to speak, and empathy (Heylen, 2006). Because head movement can have a profound impact on both the quality and perception of social interaction, Ryan's neck mechanism was redesigned to have sufficient degrees of freedom and range of motion to naturally mimic the varied movements of a human neck.

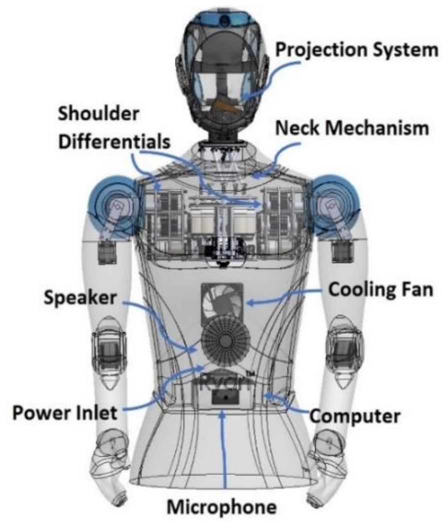
The neck mechanism of Ryan V1.0 uses a simple pan/tilt unit to produce yaw and pitch rotations of the head. This design is sufficient as a proof of concept and to verify the feasibility of the system; however, these basic head movements were perceived as unnatural by participants in the field study, thus detracting from the effectiveness of social interactions. To provide richer and more expressive interactions, the neck mechanism for Ryan V2.0 was based on a Stewart platform (Dasgupta and Mruthyunjaya, 2000) parallel manipulator with six degrees of freedom (Figure 3.3). A six-degree-of-freedom design was chosen for this version of Ryan because the human head is not limited to solely rotating about its centroid. It can translate as well (e.g., back in disbelief), and the center of rotation of the head can also be shifted along the cervical spine to produce rotations about the centroid of the head or the base of the neck for different effects. Therefore, a purely rotational mechanism is not sufficient to mimic the sophisticated head movements that play a vital role in emotive dialog.

Projection Enhancements: Ryan uses character animation technologies to project life-like 3D models onto a translucent mask to display the rich natural speech and facial expressions. This enables it to display a variety of different characters to suit the preferences of the individual user with minimal effort and create highly dynamic facial expressions without mechanical actuators and components that are prone to failure.

The Ryan V1.0 uses a bulky projector positioned outside of the head that was replaced with a far smaller form factor projector along with a mirror and lens assembly that fits fully and compactly inside Ryan's head enclosure. This setup looks more aesthetically pleasing and natural. It also protects sensitive components from damage. Figure 3.4 depicts the new design and the placement of the internal components.



(a) Ryan V2.0



(b) Ryan V2.0 internals

Figure 3.4: (a) frontal view of full Ryan CompanionBot design and (b) transparent view illustrating positioning of key components and sensors.

3.2.2 Software

For the second version of Ryan, the software stack was rewritten from scratch. In the new design, every conceptual module is represented by a ROS(Quigley et al., 2009) node. Instead of using Realsense SDK for emotion recognition, I created a facial expression recognition model based on MobileNet (Sandler et al., 2018) deep neural network architecture trained on AffectNet (Mollahosseini et al., 2017) dataset. Kaldi (Povey et al., 2011) was used for speech recognition and CereProc (Garrido et al., 2008) for text to speech. We created a "cartoonish" 3d face model (see Figure 3.5) that fits the new body design better. A few new games were added to Ryan and the tablet UI was redesigned.

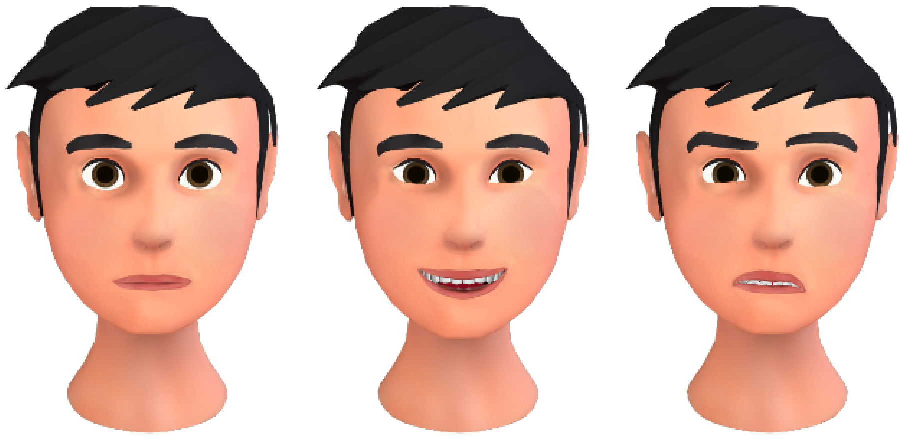


Figure 3.5: A model with different facial expressions designed for Ryan.

3.3 Ryan V3.0

In version 3.0 of Ryan, we introduce new features and upgrade the hardware to make Ryan more robust and aesthetically pleasing.

3.3.1 Hardware

The microphone in Ryan is upgraded to a mic array, allowing Ryan to filter environmental noise and improve the accuracy of text transcription. The complex Stewart platform neck introduced in V2.0 is replaced with a simpler 3 degree-of-freedom neck. The arms are upgraded and other than the proximity sensors in the base, now the arms also monitor their torque to detect collision. An NFC reader is added to Ryan to make it possible to have multiple users log into the same robot using a unique NFC tag.

3.3.2 Software

A new authentication backend is designed and implemented that enables Ryan to support user profiles. This also enables us to collect each users' analytics automatically. A web

UI is designed and created that visualizes the recorded data for each user. The offline Kaldi speech recognition software is replaced by an online service by Microsoft, to improve the accuracy. A new dialog manager software based on Facebook's ParlAI system is created for generative and unbounded conversations. Finally many new games, yoga, Spotify, more custom faces and an in-house text-to-speech is also integrated into Ryan V3.0. Figure 3.6 shows the latest design and some of the features of the latest version of Ryan.

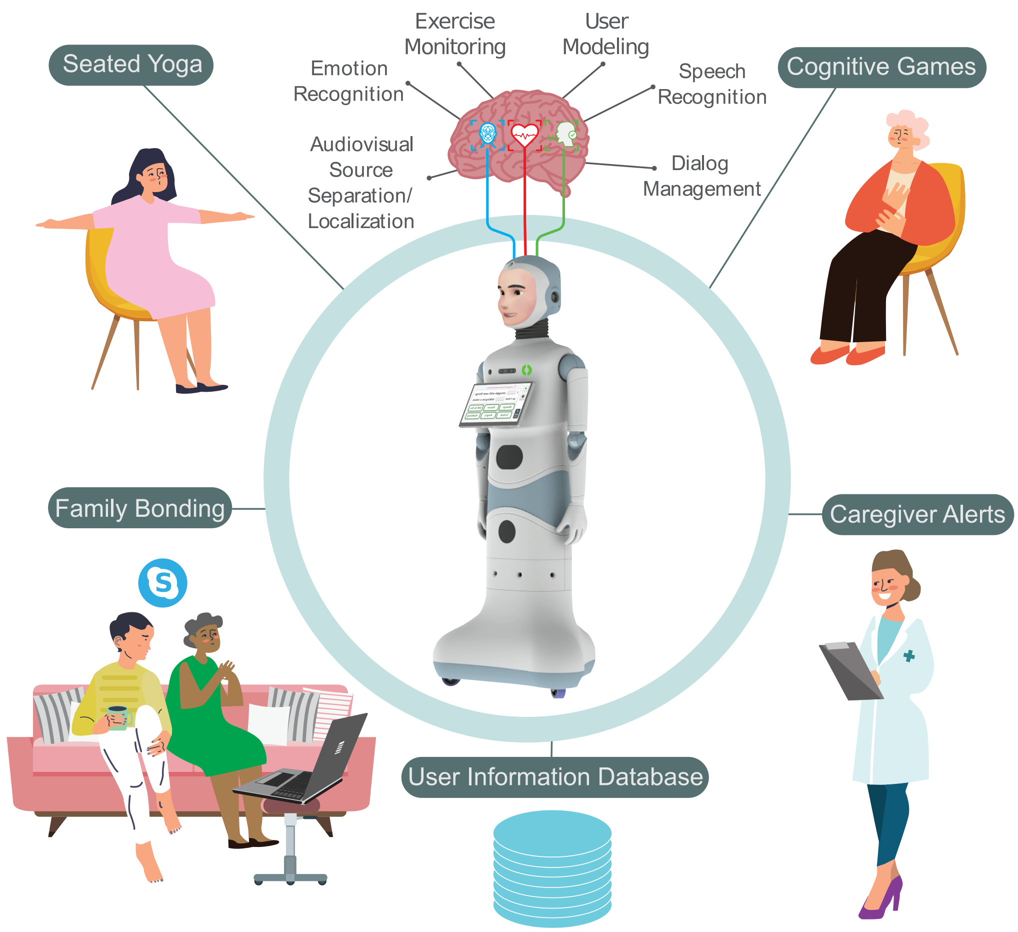


Figure 3.6: An overview of the Ryan's latest design and features (Ryan V3.0).

3.4 Summary

Table 3.1 contains a summary of some of the changes made to Ryan over the years. The changes made to Ryan are the results of our field studies and the user feedback. Ryan v3.0 is more expressive, aesthetically pleasing, stable, feature rich, and extensible.

Table 3.1: A summary of some of the improvements between Ryan version 1.0 and version 3.0 and the reasoning behind them. Ryan V2.0 was not used in any experiments and was refined into Ryan V3.0.

	Feature	Ryan V1.0	Ryan V3.0	Reason for change
Hardware	Arms	Passive	Active	To encourage the users to do more physical activities. Active arms enable Ryan to teach chair yoga and makes Ryan more expressive.
	Tablet	Vertical	Horizontal	We changed the orientations to be able to show larger pictures and videos. The angle of the tablet is also adjustable, this helped with the viewing angle and ergonomics of using Ryan.
	Projector	Outside the head	Enclosed in the head	This was done to improve Ryan aesthetically.
	Body Shell	Thermoformed	3D printed	3D printing allowed us to change the shape of Ryan's body to make it more aesthetically pleasing.
	RGBD Camera	MS Kinect	Intel Realsense	Switching to a smaller and more generic RGBD camera allowed us to use a more accurate skeleton tracking algorithm for Ryan's game and yoga.
Software	Software Framework	.Net state machine	ROS	Robot Operating System (ROS) is the standard framework used in robotics. Switching to ROS simplified the integration of the active arms with the rest of the system.
	Facial Animation	MS XNA	Unity3D	Unity3D is a an advance game engine that allows for better animations and interoperability with other modules.
	Speech Synthesis	MS TTS	Custom DNN model	Using a voice actor and training a DNN model improved the quality of synthesized audio and made Ryan's voice more natural.
	Visual Speech	Generated by a speech recognition toolkit	Generated by the Speech Synthesis software	This improved the accuracy of the visual speech and lip movement and lower the computational complexity of the system.

Continued on the next page

Table 3.1 – Continued from the previous page

Feature	Ryan V1.0	Ryan V3.0	Reason for change
Speech Recognition	Windows STT	Azure Cognitive Services	The state-of-the-art speech recognition offered by Microsoft Azure reduced the number of errors in the speech transcription and in turn improved the quality of the conversations
Facial Expression Recognition	Kinect SDK	Custom DNN model	Using a more accurate facial expression recognition is crucial in a emotionally intelligent robot.
User Authentication	N/A	Custom Node.js backend	The first version of Ryan did not have any user authentication ability. Adding this feature allows multiple users to use the same robot while preserving their data and privacy.
Music	Provided by the user	Spotify	This improvement allows the user to have access to any song they would like to listen to. Playing music from a specific decade was also a user requested feature.
Dialog Manager	ProgramR	KatieBot	ProgramR is a retrieval-based chatbot which requires a corpus of pre-written dialog. KatieBot is a hybrid between ProgramR and BlenderBot from Facebook. BlenderBot is a generative model. This allows Ryan to have sensible conversations about virtually any topic.

Chapter 4

Ryan: Human-Robot-Interaction Studies

4.1 Introduction

In this chapter two preliminary studies are presented. These studies have been conducted in the lab to evaluate the user perception of Ryan as a social robot, the effects of the embodiment of Ryan, and finally the feasibility of having an empathic social robot.

4.2 Ryan's embodiment

In this study, my colleague and I aimed to explore how the unique features of Ryan influence three major elements of human-robot face-to-face communication, namely the perception of visual speech, facial expression, and eye gaze. The details and results of the visual speech and facial expression studies are presented in Mollahosseini et al. (2018b). In this study, I focus on the eye gaze. Below I present the findings of my study.

4.2.1 Eye Gaze

Eye gaze is one of the most basic and important features of the human face for non-verbal communication. Humans incorporate gaze both consciously and unconsciously into various human-human interaction schemes (Chen and Yeh, 2012). For example, neurons in the primate visual cortex can respond selectively to eye gaze, head orientation, or even a combination of both (Perrett et al., 1985). Eye gaze serves several different functions such as capturing attention, maintaining engagement (Cassell, 2000), conveying information about emotional and mental state (Ruhland et al., 2014), augmenting verbal communication (Emery, 2000), orchestrating turn-taking and deictic reference (Kendon, 1967).

Considering the importance of eye gaze in social interaction, it is not surprising that social gaze behavior has been studied on many robotic platforms (Imai et al., 2002; Mutlu et al., 2009; Yoshikawa et al., 2006). Mechanical and Android robotic platforms control eye gaze by using actuators in the eyeballs. However, these actuators may not be fast or accurate enough to replicate the movement of human eyes. The movement of the human eye is controlled by three pairs of muscles and it can reach an angular speed of about 400°/sec with a time of 200ms to initiate (Pateromichelakis et al., 2014). Computer graphics animations, on the other hand, have a greater capability to produce a natural-looking eye gaze (Cassell, 2000; Ruhland et al., 2014). However, it is known that the perception of 3D objects that are displayed on 2D surfaces is influenced by the Mona Lisa effect (Todorović, 2006). Hence, the lack of physical embodiment and physical presence may constrain the perception of virtual agents' eye gaze.

4.2.2 Related Work

Many studies in vision science have evaluated head-eye gaze, but only on telepresent faces (Allison et al., 2000; Baron-Cohen et al., 1995; Itier and Batty, 2009; Sweeny et al.,

Table 4.1: Summary and overview of literature comparing perception of eye gaze in different conditions.

Work	Agent	Condition*				EG†	Description	Results**
		CR	TR	VA	GT			
Anstis et al. (1969)	TV		✓		✓	✓	• A horizontal scale (ruler) was used • Video of a human used for TR • The agent's head was rotated with -30°,0° and 30° angles	• Errors were greatest when head rotation and eye rotation were incongruent.
Delaunay et al. (2010)	LightHead	✓	✓		✓	✓	• A grid with 100 cells was used • Video of a human used for TR • Instead of head rotation, subjects viewed the Agent with 0° and 45° angles	• CR performed better than TR • GT performed significantly better than other conditions, in both frontal and side view situations
Al Moubayed and Skantze (2012)	Furhat	✓			✓		• A grid with nine cells was used • Vergence, parallel eyes, static and dynamic eyelids	• Perception of gaze was significantly worse when the head was moving compared with eye movement alone. • No significant difference between gaze with and without vergence.
Moubayed et al. (2012)	Furhat	✓		✓			• Mona Lisa effect studied on five subjects sitting around a circle. • Only eye rotation studied	• Gaze was perceived more accurately on CR
Misawa et al. (2012)	LiveMask	✓		✓			• Photos of a person looking from -30° to 30° • Instead of rotating the head, subjects' view angle was changed	• CR was significantly better than VA • The Mona Lisa effect occurred in VR
Mollahosseini et al. (2014b)	Expressionbot	✓		✓			• Mona Lisa effect studied on five subjects sitting around a circle	• Discrimination of eye gaze was better on CR
This work	Ryan	✓	✓	✓	✓	✓		

* CR, TR, VA, and GT stand for Copresent Robot, Telepresent Robot, Virtual Agent, and Ground Truth (human) respectively.
†EG stands for Emergent Gaze which is defined as simultaneous movement of head and eye-gaze.
** Only the relevant finding from the original papers are reported in this summary.

2012). Although embodiment and presence have been studied individually, there is not a comprehensive study that distinguishes the role of embodiment and presence in gaze perception. Gaze perception of a physically present human agent and his video was studied on a TV set by Anstis et al. (1969). In this classic study, subjects were asked to report the point on a glass screen at which the agent (TV or a human) was looking. To simulate head rotation in the telepresent condition, the TV set was rotated. The agent's head was rotated at -30°, 0° and 30° angles. The study found that eye gaze was much better perceived on a physically present human agent than on its telepresent counterpart, and the perception of gaze was distorted with the rotation of the TV.

Delaunay et al. (2010) studied gaze perception on the LightHead robotic face, its telepresence, and the gaze of a human agent. A vertical glass screen with a 10x10 grid was placed between the agents and the subjects, and subjects were asked to report the gaze point when viewed from a frontal and 45° angle. Since asking a human to hold his/her head steady in a 45° position was not possible and chin/forehead rests did not allow horizontal rotations, to study the effect of head rotation, subjects were instead moved to a position with a 45° angle with respect to the agent. Under these conditions, subjects judged gaze from the video and the robot in both frontal and 45° view situations with equal sensitivity.

Al Moubayed and Skantze (2012) compared the perception of eye gaze on Furhat robotic face with a human agent under different conditions (i.e. presence of vergence, static/dynamic eyelids, etc.). They took a different approach by asking the agents to look at nine points on a table between the agent and the subjects. In this case, there was no significant difference between gaze with vergence and without vergence. Furthermore, head movement appeared to be more effective in influencing judgments along the horizontal axis, while eye movement dominated judgments along the vertical axis. Regardless of the conditions, the gaze of the human agent was perceived better than the gaze of the robot.

Studies show that virtual agents suffer from the Mona Lisa effect (Misawa et al., 2012; Mollahosseini et al., 2014b; Moubayed et al., 2012), in which the eyes in a picture appear to be looking at the viewer regardless of their location in front of the picture. For example, Moubayed et al. (2012) studied the Mona Lisa effect on a virtual agent and its 3D projection on Furhat robotic face. Five subjects were simultaneously seated around the agent, each of whom was asked to report their perception of the agents' eye gaze direction. The results showed a clear Mona Lisa effect in the virtual agent since many subjects perceived a mutual gaze at the same time.

Table 4.1 summarizes several studies on eye gaze perception and their most relevant findings. The majority of these studies report that physical presence plays a greater role in perception of an agent's eye gaze than physical embodiment. Presumably, having a 3D view of the nose direction, the eye position, and their composition help viewers to perceive eye gaze direction more accurately. In addition, few studies have explored emergent gaze. Emergent gaze occurs when the visual system integrates global information about the rotation of the head with local information about eye rotation, to compute a distinct metric of gaze present in neither feature alone (Cline, 1967; Kinya and Mitsuo, 1984; Kluttz et al., 2009; Langton et al., 2004; Otsuka et al., 2014; Sweeny and Whitney, 2017; Wollaston, 1824). This approach to measuring gaze perception has been surprisingly underutilized in robotics work.

The present study evaluates the perception of emergent gaze, while at the same time comparing the roles of embodiment and presence of the robot. One of the reasons that emergent gaze has not been studied extensively both with humans and robots is the difficulty inherent in controlling the movements of a human agent. Rotating a human's head and eyes to an exact position requires special apparatuses and complicates the experiment process. Hence, most studies of gaze either do not include a condition with a human agent, or they use a typical chin/forehead rest to fix the human's head in place, which precludes examination of emergent gaze.

4.2.3 Methodology

To evaluate the accuracy of agents' eye gaze in the current investigation, the agent looked at a particular point on a glass divider located between the agent and the subjects. A horizontal line with fifty-one equidistant points was drawn on the glass. The agent looked at a point on the glass screen, and subjects were asked to report their perception of the agent's gaze direction.

In order to precisely set eye gaze toward a target point, we needed to rotate the agents' eyeballs such that the pupils were directed towards the target point. In this study, the target points were at agent's eye level; hence we only needed to change the yaw angle for the eyes. Assuming the face is frontal (rotated zero degrees), the yaw angle for the right and left eyes (α_r and α_l, respectively) is calculated as:

$$\alpha_r = \frac{\pi}{2} - \arctan \frac{x + E_r}{D_r} \tag{4.1}$$

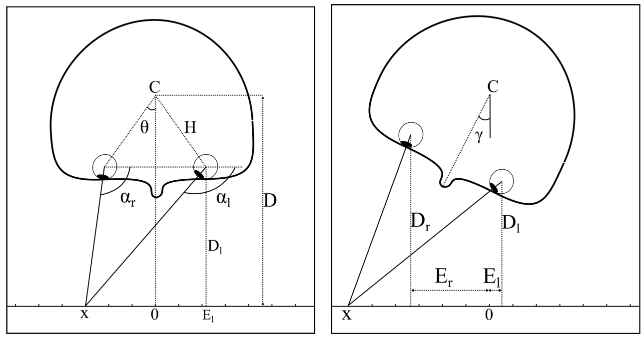
$$\alpha_l = \frac{\pi}{2} - \arctan \frac{x - E_l}{D_l} \tag{4.2}$$

where $x \in [-75cm, 75cm]$ is the target point on the glass screen. E_r and E_l are the distance of the right and left eye from the center of the glass screen on the x-axis, and D_r and D_l are the distance of the right and left eyes from the glass screen on the y-axis, calculated as:

$$E_r = E_l = H \times \sin(\theta) \tag{4.3}$$

$$D_r = D_l = D + H \times \cos(\theta) \tag{4.4}$$

where H is the distance of the head pivot point (C) from the center of the eyes, θ is the angle between the eyes and the head pivot point, D is the distance of the head pivot point to the glass screen. Figure 4.1a shows the schema and the variables used in these calculations.



(a) Head facing forward (b) Head rotated by γ

Figure 4.1: Schema and the variables used in the calculating eye gaze angle (Drawing not to scale).

When the head is straight, $D_l = D_r$ and $E_r = E_l$. If the head is rotated by $\gamma°$ (Figure 4.1b), the values of E_r and D_r in Equations (4.1) and (4.2) are changed as follows:

$$E_r = H \times \sin(\theta + \gamma) \tag{4.5}$$

$$E_l = H \times \sin(\theta - \gamma) \tag{4.6}$$

$$D_r = D - H \times \cos(\theta + \gamma) \tag{4.7}$$

$$D_l = D - H \times \cos(\theta - \gamma) \tag{4.8}$$

In the above equations, we assumed that the agent does not have any facial curvature in the eye area (Figure 4.2-left). If the face has an angle (ϵ) in the eye area (Figure 4.2-right), Equations (4.1) and (4.2) will change as follows:

$$\alpha_r = \frac{\pi}{2} - \arctan\frac{x + E_r}{D_r} - \epsilon \tag{4.9}$$

$$\alpha_l = \frac{\pi}{2} - \arctan\frac{x - E_l}{D_l} + \epsilon \tag{4.10}$$

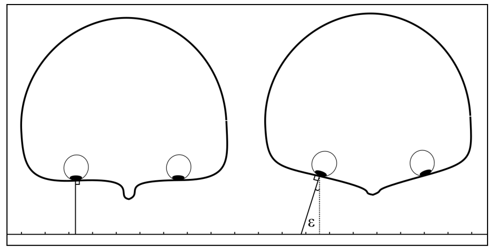


Figure 4.2: Mask with flat eye region (left) and with angled eye region (right).

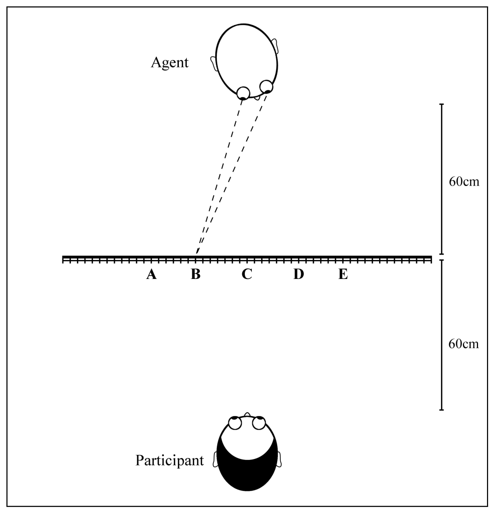
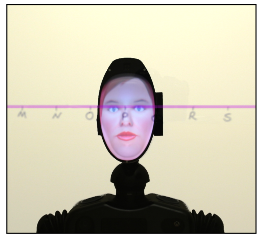


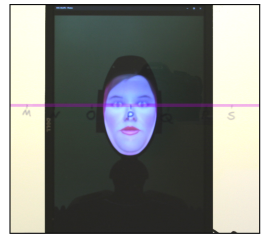
Figure 4.3: Perception of eye-gaze setup. Fifty-one points with three centimeters distance from each other were marked on the glass. The agents looked at only A, B, C, D, and E points located at -39, -21, 0, 21 and 39 centimeters from the center respectively.

4.2.4 Eye Gaze Experiment

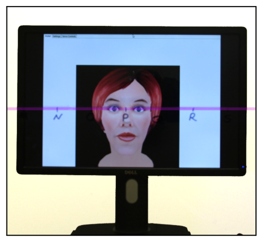
We examine the perception of eye gaze with 23 subjects, 7 women and 16 men, with an age range of 21-40 years (mean = 28.4, SD = 5.5), each of whom had normal or corrected to normal vision. To evaluate the role of embodiment and presence in the perception of the agent's eye gaze, four conditions (VA, CR, TR, and GT) were examined in this experiment. In each condition, the agent looked at a particular point on a glass divider located between the agent and the subjects. The subjects were then asked to report their perception of where the agent was looking.



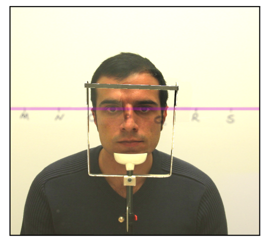
(a) Copresent Robot



(b) Telepresent Robot



(c) Virtual Agent



(d) Ground-Truth

Figure 4.4: Eye gaze different conditions.

The subjects were seated in front of the glass screen and then asked to keep their head still on a chin-forehead rest and look straight at the agent at a distance of 120 cm. To simulate the most accurate head rotation and avoid a Mona Lisa effect, which is common when viewing a face on a flat screen, in the VA condition we presented rotations of the animated head itself rather than rotations of the screen portraying the head. Figure 4.3 illustrates the eye gaze evaluation setup.

Fifty-one points, three centimeters apart, were marked by letters and numbers on the glass. However, the agents looked at only five points located at -39, -21, 0, 21 and 39 centimeters (with zero as the middle point of the glass divider). Hereafter, these points are referred to as *A, B, C, D* and *E*, respectively (shown in Fig. 4.3). Subjects were not aware of the agent's restricted gaze targets, and they were instructed that the agent may look at any point on the glass. Figure 4.4 shows photos of different conditions viewed from the subject's position.

We examine the emergent perception of eye gaze (i.e., the integration of head rotation information with eye position). In particular, there were five possible head rotations (-30°, -16°, 0°, 16°, and 30°), and in each head position, the eyes were shifted toward the five points on the glass screen. An example of this condition is shown in Fig. 4.3, where the agent's head is rotated toward +16° and the eyes are directed at point *B*.

The method described in Section 4.2.3 was used to calculate the angle for the agent's eyes in CR and TR scenarios. The dimensions of the robot head for CR and the 3D model for VA were measured, and depending on the target point on the glass screen, the eyes of the robot/3D model were rotated toward the target point. The measurement used in CR was: $D = 73cm$, $H = 13.35cm$, $\theta = 13°$, and the measurement used in VA was: $D = 70cm$, $H = 10.45cm$, $\theta = 17°$. Since a mask with a flat eye region was used in CR and a flat screen was used in VA, the value of ϵ was set to $0°$.

A Canon EOS 80D DSLR camera was used to take pictures of the robot from the point of view of the subject. The captured images were calibrated to the size of the robot head. Using this method, from the point of view of the subject, the agent in both CR and TR had the same size and proportions and, in theory, the same direction of eye gaze (if we took a picture from the subject's point of view, it would look the same). The difference was that the TR condition featured a 2D representation of the CR condition.

To keep the human agent's head in an exact head rotation angle consistently during the GT experiments, we modified a chin/forehead rest to rotate and then stabilize in 1° increments. In the GT condition, a human was seated in the place of the agent and looked at the points on the glass, while keeping his head still on this chin forehead rest and his shoulders facing directly forward.

In all four conditions, first, the agent's head was rotated to one of the five angles (-30°, -16°, 0°, 16°, and 30°) randomly. Then at each of these head angles, the eyes were rotated to gaze at one of the 10 points on the board (two trials for the five targets *A, B, C, D* and *E*) randomly. The subject was asked to close his/her eyes between each trial to eliminate any effect of seeing the agent adjust his head and eyes. In total, each subject reported 50 gaze directions (5 angles $\times$ 5 points $\times$ 2 trials) for each condition. Each condition was run in a block lasting five minutes, and the subjects were asked to leave the room for two minutes until the room was set for the next condition.

Four different agent conditions (VA, TR, CR and GT) were presented in random order to the subjects, and subjects were asked to report their perception of the point at which the agent was looking. The accuracy was calculated by measuring the error in each subject's reports of eye gaze. The gaze perception error was defined as the absolute distance between the point that the subjects reported and the actual target point at which the agent was looking.

4.2.5 Eye Gaze Results

We performed a 5 (head rotation) $\times$ 5 (eye gaze) $\times$ 4 (agent conditions: CR, TR, VA and GT) ANOVA with agent condition, head rotation, and target point as within-subject factors. The dependent variable was the gaze perception error. This analysis revealed a significant main effect of agent condition [$F(3, 66)$ = 134.55, p <.0001]. We also found the main effects of head rotation [$F(4, 88)$ = 70.25, p <.0001] and eye gaze [$F(4, 88)$ = 31.39, p <.0001]. This analysis also revealed an interaction between agent condition and head rotation [$F(12, 264)$ = 11.17, p <.0001], but the interaction between agent condition and eye gaze was not significant [$F(12, 264)$ = 95.16, *n.s*]. Figure 4.6 shows the estimated marginal means of gaze perception error for different agents, head rotation angle and target points. As shown, the differences between the agent conditions depended on head rotation, but not eye gaze.

Table 4.2: Average and proportional error with respect to human ground-truth for different agent conditions.

	Average Error ± STD (cm)	Proportional Error to GT
GT	7.88 ± 2.90	-
CR	10.50 ± 3.11	33.26%
TR	11.04 ± 3.16	46.47%
VA	13.04 ± 2.88	65.57%

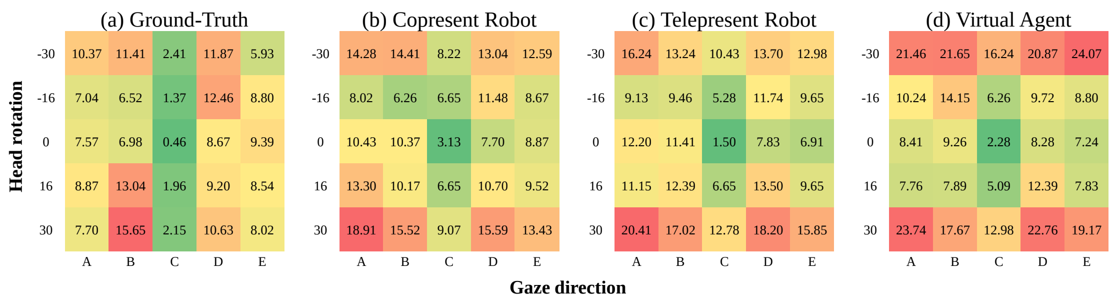
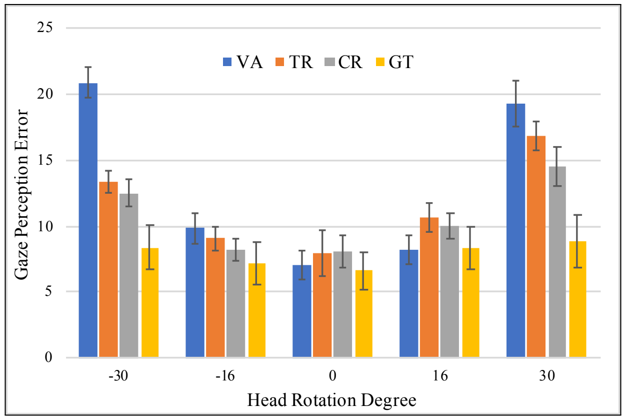


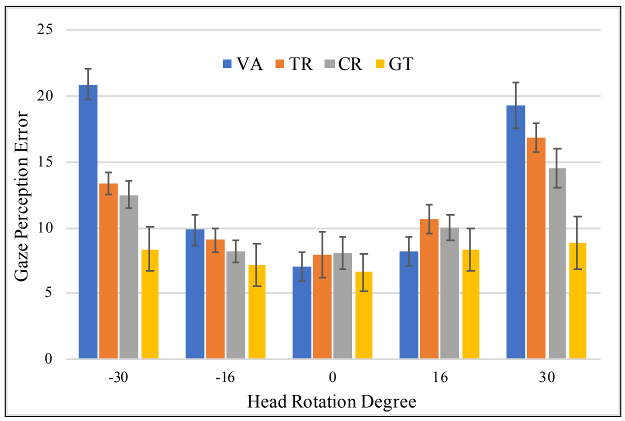
Figure 4.5: Average absolute error of gaze perception in different conditions [best viewed in color].

Table 4.2 shows the average and standard deviation of error for each condition and proportional error with respect to human ground truth. The results indicate that the eye gaze was better perceived in CR than in TR and VA, with 13.21% and 32.23% lower proportional errors, respectively. Figure 4.5 shows the average error (cm) in the perception of different agents' eye gaze for different head rotation and target points. As Fig. 4.5-(a) shows, when the eye gaze was directly toward the subject's face (point C), the perception of eye gaze had a relatively negligible amount of error. In other words, the subjects were able to recognize mutual eye contact with high precision in the human agent. The same pattern emerged in the CR and TR conditions. Interestingly, subjects discriminated mutual eye gaze poorly in the VA condition, especially with incongruent head and eye rotations.

In particular, when the head was rotated to its extremes (-30° and 30°), perception of gazes directed toward points *B* and *D* had higher error than gazes directed toward points *A* and *E*. This suggests that subjects had difficulty accurately recognizing the direction of



(a) Head rotation



(b) Target Point

Figure 4.6: Estimated marginal means of gaze perception error for different agents and (a) head rotation angles and (b) different gaze target points. The target points A, B, C, D correspond to -39, -21, 0, 21 and 39cm from the center, respectively.

gaze when the rotation of the head was incongruent with that of the eyes. Hence, subjects may have guessed a point at the far end of the glass screen, giving them more room for error at points *B* and *D*.

As shown in Fig. 4.5, eye gaze of the virtual agent was seen with a notable amount of error ($\sim$24cm) when combined with a strong head rotation. This could be because the animation lacked binocular depth cues because it was present on a flat screen. This could have made perception of head rotation more difficult, while the embodiment of the robot helped subjects to recognize the head angle better.

In order to more directly measure the effect of agents' embodiment and presence, we removed human GT from the analysis and performed a 5 (head rotation) $\times$ 5 (eye gaze) $\times$ 3 (agent conditions: CR, TR, VA) ANOVA with agent condition, head rotation, and eye gaze as within-subject factors. This analysis revealed the main effects of agent [$F(2, 44)$ = 8.740, p = .001], head rotation [$F(4, 88)$ = 64.95, p <.001] and eye gaze [$F(4, 88)$ = 16.39, p <.0001]. Similarly to the previous analysis, and as shown in Figure 4.6, there was a significant interaction between agent condition and head rotation [$F(8, 176)$ = 8.75, p <.0001], but the interaction between the agent condition and eye gaze was not significant [$F(8, 176)$ = 23.98, *n.s*].

Since there was an interaction between the agent condition and head rotation, we performed pairwise two-tailed t-test comparisons between agent conditions at different head rotations. Table 4.3 shows pairwise p-value and Cohen's d effect-size between agent conditions. As shown, the embodiment (Research Question 1) improved the perception of eye gaze at -30° and 30°, as indexed by significant differences between the TR and VA conditions ($p < .001$ and $p = .023$ with large effect sizes $d = 1.22$ and $d = 0.69$ respectively). Physical presence did not improve the perception of eye gaze (Research Question 2), as the differences between the TR and CR conditions were not significant at any head angle. There were also significant differences between CR and VA at -30° and 30°, both $p < .001$ with large effect sizes $d = 1.49$ and $d = 0.89$ respectively (Research Question 3). Because TR and VA were both significantly different at these head angles, we conclude that improvement in the perception of eye gaze compared to CR is mainly due to embodiment rather than presence of the robot. And in particular, the embodiment of the robot highly affected the precision of gaze perception combined with extreme head rotations in a frontal situated setting.

These findings are consistent with previous studies showing that the perception of a robot's eye gaze is more accurate than that of a virtual agent (Misawa et al., 2012; Molla-

Table 4.3: Pairwise comparison (LSD p-value) and Cohen's d effect size of users' perception of eye gaze at different head rotations. Significant pairs are shown in bold.

	TR vs CR		VA vs CR		VA vs TR	
Head Angle	p	d	p	d	p	d
-30°	.660	0.13	**<.001**	**1.49**	**<.001**	**1.22**
-16°	.479	0.21	.190	0.39	.5484	0.17
0°	.890	0.04	.278	0.32	.269	0.32
16°	.599	0.15	.116	0.47	.158	0.42
30°	.217	0.36	**.004**	**0.89**	**.023**	**0.69**

hosseini et al., 2014b; Moubayed et al., 2012). There was no difference in gaze perception when seen on a robotic agent or its telepresence, which is consistent with a study by Delaunay et al. (2010). We also did not observe a significant difference between gaze perception on the telepresent robot and virtual agents—a comparison which has not been addressed in previous studies.

4.3 Incorporating affection in spoken dialogue in a social robot

In this section, I present our effort on incorporating an automated Facial Expression Recognition (FER) system based on deep neural networks into the spoken dialogue of a social robot (Ryan) to extend and enrich its capabilities beyond spoken dialog and integrate the user's affect state into the robot's responses. The results of this study are published in Mollahosseini et al. (2018a). Here, I present the details of my efforts in creating the facial expression and dialog modules used in this study. In order to evaluate whether the incorporation of FER in spoken dialogue can improve the social capabilities of Ryan, we conducted a series of HRI experiments. In these experiments, the subjects watched some videos and Ryan engaged them in a conversation driven by user's facial expressions perceived by the robot. I measured the accuracy of the automated FER system on the robot when interacting with different human subjects as well as three social/interactive aspects, namely task engagement, empathy, and likability of the robot. The results of this HRI study indicate that the subjects rated empathy and likability of the affect-aware Ryan significantly higher than non-empathic (the control condition) Ryan. Interestingly, we found that the accuracy of the FER system is not a limiting factor, as subjects rated the affect-aware agent equipped with a low-accuracy FER system as empathic and likable as when facial expression was recognized by a human observer.

4.3.1 Automated FER System

Facial expression plays a vital role in social interaction and is one of the most important nonverbal channels of recognizing humans' internal emotions. Numerous computer vision and machine learning algorithms have been proposed in the literature for automated Facial Expression Recognition (FER) (Tian et al., 2011). The majority of these techniques are based on supervised machine learning methodologies that require annotated samples for training, and their performance highly depends on extracted features from the samples and the amount and diversity of annotated training samples. Several available FER systems are trained on databases containing posed expressions acquired in a controlled lab environment with limited numbers of subjects and few samples per expression. Therefore, these systems lack sufficient generality when used in an uncontrolled HRI system.

Recently, databases of facial expression and affect in the wild have received much attention (Goodfellow et al., 2015; Mollahosseini et al., 2016). In this work, we use a newly released database of facial *Affect* from the Inter*Net* (called *AffectNet*) which is publicly available to the research community (Mollahosseini et al., 2017). AffectNet contains more than 1M images with faces and extracted landmark points. The database is created by querying different search engines using emotion-related tags in six different languages. Twelve human experts manually annotated 440,000 of these images in eleven discrete categories (i.e., Neutral, Happy, Sad, Surprise, Fear, Anger, Disgust, Contempt, None, Uncertain, and Non-face) and dimensional model of affect (i.e., valence and arousal).

Since we only study four facial expressions in this work, we trained a 50-layer Residual Network (ResNet) (He et al., 2016) in five classes of neutral, happy, surprise, sad, and disgust of affectNet database. The ResNet architecture is a state-of-the-art CNN with added shortcut connections, i.e., a linear transform of each layer's input to the layer's output. Adding the shortcut connection eases the training of deeper networks (more than 100 layers) and prevents degradation problem (the phenomenon that the accuracy becomes saturated and then degrades rapidly (He et al., 2016)). The residual connection has produced state-of-the-art performance in several computer vision applications such as visual object detection (He et al., 2016), audio classification (Hershey et al., 2017), and facial expression recognition (Hasani and Mahoor, 2017).

During the experiments, subjects' faces were captured by a webcam installed on the video player monitor. The OpenCV face recognition library was used to detect faces in the images, and 66 landmark points were found using a face alignment algorithm using local binary regression features (Ren et al., 2014; Yu, 2016). We used these points to register faces to an average face using an affine transformation. Once the faces have been registered, the face regions were cropped, resized to 48$\times$48 pixels, and fed into the trained network for classification.

We used a K40 GPU for training the network and an Intel Core i7 CPU during inference. Face detection, registration and expression classification take $\sim$20ms, enabling us to process five frames per second. A majority voting is used to determine the user's facial expression while watching the video. As videos trigger emotions in a few scenes and users had neutral faces in the rest of time of watching the videos, the frames with emotions detected as neutral faces were discarded by the probability of 0.5 in a majority voting scheme.

4.3.2 Empathic Conversations

According to Preston and De Waal (2002) empathy reaction can be a function of three factors:

1. Be affected by and share the emotional state of another.

2. Assess the reasons for emotional state.

3. Identify and adopt other perspectives.

Taking into account these elements and the previously given definition of empathy, we propose the following features that need to be embodied in our empathic robot:

- The robot should be capable of recognizing, understanding, and interpreting the user's emotional state (facial expression in this experiment).

- The robot should be capable of expressing its emotion using both verbal and non-verbal cues.

- The robot should be capable of taking perspective, being supportive and have self-correction to adopt other perspectives.

The robot recognizes the user's facial expression while watching the videos. Based on the affective state of the user, the robot appraises the situation and generates empathic responses, e.g., "congruent facial expressions" in tune with the user's affective state, "perspective taking", "being supportive", and "self-correction".

A set of predefined empathic responses based on the perceived affect state and conversation with users was carefully designed. Figure 4.7 shows an example of empathic

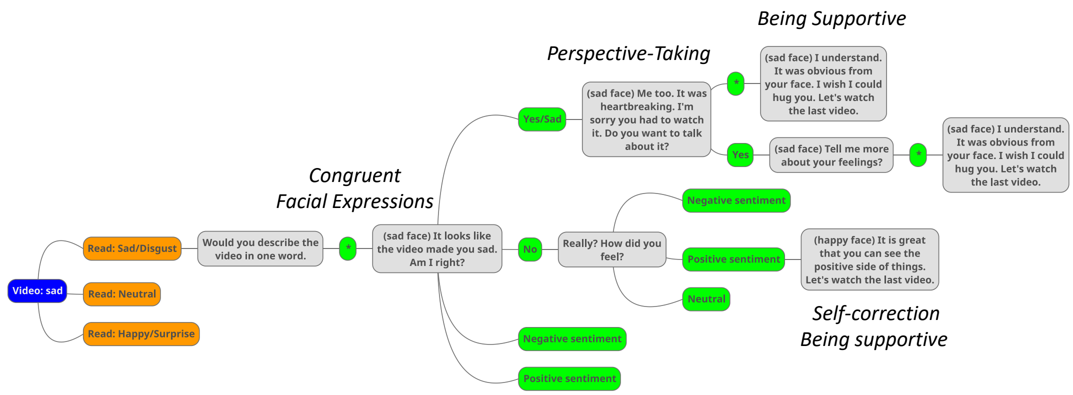


Figure 4.7: Example of empathic conversation map after showing a video intended to elicit a sad emotion.

conversation map after showing videos intended to elicit sad emotion. As shown, if Ryan recognizes sadness, she shows a sad face [congruent facial expressions] and says *"It looks like the video made you sad. Am I right?"*. If the user confirms that he/she was sad, the robot keeps the sad face and says *"Me too. It was heartbreaking. I am sorry you had to watch it."* [perspective- taking], *"Do you want to talk about it?"*. Based on user's response, the robot will say *"I understand. It was obvious from your face."* [perspective- taking], *"I wish I could hug you"* [being supportive]. If the user did not have a negative effect (e.g., the user had a neutral face) and the robot recognized it incorrectly, the robot stops showing a sad face and says *"Oh. Seems like I misinterpreted your face"* [self-correction], *"You are focused on the task. Good job"* [being supportive]. Refer to Mollahosseini et al. (2018a) for the results of this study analyzed by Dr. Ali Mollahosseini.

Chapter 5

Studying Ryan as a Socially Assistive Robot for Older Adults

5.1 Introduction

Developing and studying robots as an assistive tool for healthcare professionals is a growing area of research due to the rapid growth in the number of elderly people and the demand for specialized caregivers. Socially Assistive Robotics (SAR) Feil-Seifer and Mataric (2005) focus on improving elderly people's quality of life, mental health, and socioemotional well-being. Social robots are used as companions Taggart et al. (2005) or therapeutic play partners Leite et al. (2010). The essential feature that defines SAR is using social interactions rather than physical interactions to help the user Rabbitt et al. (2015). The focus of this chapter is on SAR and the companionship it provides for elderly people with moderate depression and/or dementia.

Dementia is an overall term for diseases that deteriorate individuals' memory and other mental skills. Dementia can significantly reduce elderly individuals' ability to live independently and safely in their homes. It is one of the most costly diseases and requires hours of specialized care for each person Abbott (2011). Associated with the decline in cognitive abilities, depression is one of the symptoms of dementia Marti et al. (2006).

Thus, there is a critical and growing demand in the community to find effective ways to provide care for elderly people with dementia. There is an emerging research field in robotics that aims to use social robots to engage effectively in social and conversational interaction with elderly individuals with dementia to improve their socioemotional behaviors, cognitive functions, and well being. We conducted a pilot study to demonstrate the feasibility of using Ryan Companionbot, a perceptive and empathic conversational humanoid robot, to improve the quality of life of elderly individuals with moderate dementia and/or depression. In this study, we are using spoken dialog combined with a long list of other stimuli, such as eye gaze, head movement, and facial expressions, as the primary form of communication between the subject and the robot. Specifically, the objective of this study is to evaluate the following fundamental research questions.

1. **Long-Term Companionship:** Would enriching the robot with a number of different features keep the subjects engaged for a long period of time?

2. **Likability and Acceptance:** Is interacting with SAR enjoyable for elderly individuals and do they accept a robot as a companion?

3. **Robot Features:** Do the results of the pilot study show that each individual looked for different features (e.g., spoken dialog system, cognitive games, family photo album narration, music playing, etc.) in the robot?

The remainder of this chapter is organized as follows. Section 5.2 reviews related work on SAR and employing social robots in elder care. Section 5.3 explains the setting of the experiment and the methodology of our pilot study to evaluate the above fundamental research questions. Section 5.4 presents the results and analysis of the experiments. The results are categorized in four subsections: long-term companionship, likability and acceptance, caregivers' feedback, and robot features. Finally, Section 5.5 concludes the chapter.

5.2 Related Work

Using SAR to help elderly individuals has recently become more relevant due to the increase in the number of elderly people, the decrease in the cost of technology and recent advances in artificial intelligence Leite (2015). Residents of nursing homes live alone with disabilities while in most cases their cognitive abilities are degraded due to old age or various types of dementia Kotwal et al. (2016). Studies suggest that social support for elderly individuals could improve their cognitive function Zamora-Macorra et al. (2017). Using SARs with a focus on the socialization aspect of Human-Robot Interaction (HRI) is a viable option to reduce the burden on caregivers while providing companionship to elderly people, improving their quality of life and avoiding depression and further degradation of their mental abilities.

Wada *et al*. Wada et al. (2003) used the Paro robot to study the long-term effect of social robots on residents of a senior care center. The results indicated that the elderly residents established a relationship with the robot, developed stronger social bonds among themselves, and also maintained a lower stress level. However, Paro lacks the ability to talk and communicate. It is shown that for a social robot to be accepted more easily it should be communicative Heerink et al. (2006) and must employ a form of communication with which humans are habituated Krämer et al. (2012).

Another key aspect to having a robot as a companion is continuous (uninterrupted) companionship, meaning having access to the robot at all times. Autonomy plays a crucial role in achieving uninterrupted companionship. Most studies conducted with social robots in elder care are either performed in a Wizard-Of-Oz (WOZ) manner Vardoulakis et al.

(2012), or were limited to a specific scenario Pineau et al. (2003). Vardoulakis *et al*. Vardoulakis et al. (2012) designed an experiment to study the long-term social companion for older adults. They used a WOZ method, and the subject had a robot at home for one week. But, since the robot was controlled remotely by an operator, the subject interacted with the robot for only one hour every day. Employing the WOZ method forces the subjects to use the robot at a specific time of the day for a short period which resembles visiting a friend rather than having a companion at home. Social robots such as Paro are autonomous and provide continuous companionship, but lack the ability of having a robust social interaction such as spoken dialog and an expressive face.

Deep social interaction is required when dealing with elderly individuals with dementia. Different robots such as Aibo, Paro, and Bandit have been used in studies on the care of elderly people with dementia Mordoch et al. (2013). Most of the robots used in these studies have not been built with the social aspect in mind. But to be able to communicate with elderly people with dementia and try to engage them in conversations and games, we need a robot that has been designed to accomplish these social goals. In the following section, we will introduce a robot designed to be social.

5.3 Pilot Study

To assess Ryan's feasibility as a companionbot, we conducted a pilot study with six elderly individuals with dementia and depression living in the Eaton Senior Community in Denver, Colorado Eaton. The robot was left in their home, and they had access to the robot at all times. Figure 5.1 shows a subject interacting with Ryan V1.0.

Figure 5.1: A subject interacting with the robot in her home.

5.3.1 Subjects

A group of six volunteered elderly individuals was selected for this study. The selection criteria included those elderly people who live alone, who were in the early-mild stage of dementia, and who may suffer from depression. Other selection criteria included the availability for a period of at least four weeks to house and interact with the robot. The selected subjects consented prior to participating in the study and the family members of the subjects were also informed to ensure they were aware of the study.

The Saint Louis University Mental Status (SLUMS) Examination Tariq et al. (2006) and the Patient Health Questionnaire (PHQ-9) Kroencke et al. (2001) were completed by each patient and scored by the caregiver prior to the experiment. The SLUMS, developed at the Division of Geriatric Medicine of the Saint Louis University School of Medicine, is a favorable screening tool for detecting mild cognitive impairment. The PHQ-9 contains nine questions and is a brief and useful instrument for screening, monitoring, and measur-

Table 5.1: Participants demographics, SLUMS and PHQ-9 Scores. Highlighted cells mean that the symptoms (i.e. Dementia and Depression) exist in the patient.

Sbj	Age	Gender	SLUMS Score	PHQ-9 Score	Living Resident
1	63	F	19	17	Independent
2	86	M	21	1	Independent
3	78	F	29	15	Independent
4	73	F	17	3	Assisted
5	71	F	25	7	Assisted
6*	79	F	28	16	Assisted

* Subject 6 participated 24 days since she became ill and hospitalized at the end of pilot study

ing the severity of depression. The SLUMS scores for people with high school education are interpreted as follows: 27-30: Normal, 21-26: Mild Neurocognitive Disorder, 1-20: Dementia. The PHQ-9 severity scores are mapped as follows: score 5-9: Minimal Symptoms, score 10-14: Minor depression, score 15-19: Major depression, moderately severe, score>20: Major depression, severe. Table 5.1 shows the demographics of the patients who participated in our pilot studies.

5.3.2 Method

To measure how effectively Ryan can provide companionship for elderly individuals with dementia, we conducted a one-on-one (robot vs. human) pilot study at the Eaton Senior Community Center. Three Ryan companionbots were manufactured for the study. Each subject had 24/7 access to Ryan in their rooms for a period of 4-6 weeks. The robot was left in the room of the elderly participant and he/she treated Ryan companionbot as his/her guest. To avoid any maintenance issues, the research team remotely monitored the status of the robots.

Each subject was interviewed to obtain their daily schedules, a set of photos for the album, topics of interest for YouTube video search, and a collection of favorite music and songs. Ryans were customized for each participant. They could call the robot any name they wanted according to their preferences. The participants' daily schedule, including reminders to take their medications, was set manually for each subject.

During the study, all subjects' interactions with Ryan, the facial emotion of the users, the conversations between Ryan and the participants, and the sentiment of the speech were logged. We analyzed the log files and computed a measurement to evaluate user interactions with Ryan during the pilot study.

5.4 Results

5.4.1 Long-Term Companionship

To measure whether Ryan can be a companion of elderly individuals in long-term, the conversations between Ryan and the participants were recorded over the period of the experiment. Conversations were on different topics such as sports, emotional states, technology, or other topics. Each conversation contains several dialogs between the subjects and Ryan. We defined a dialog as an exchange of one inquiry and response between the subject and Ryan. On average, subjects and Ryan had 198 (σ=49.2) dialogs per day, with the average length of 9.2 words per each dialog.

Figure 5.2 shows the average number of dialogs of all participants over the period of four weeks. Since SN6 became ill and hospitalized at the end of the pilot study, she participated only 24 days. Therefore, the average shown for the last 4 days are data from 5 subjects. The average number of dialog time series (shown in Fig. 5.2) is then smoothed using a moving average with a window size of five, due to variation between consecutive days and subject schedule. As shown, the average number of dialogs per day for all subjects

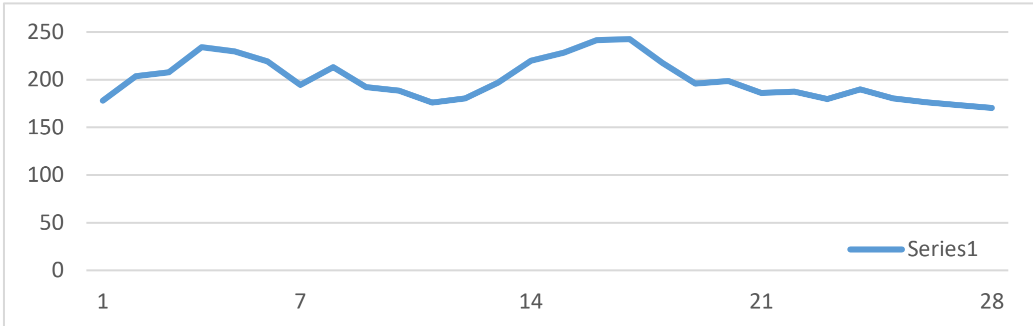


Figure 5.2: The average number of dialogs between participants and Ryan has not decayed over a period of four weeks (One subject interacted with the robot for three weeks).

did not decrease for four weeks. In other words, the subjects kept their interest in having conversations with Ryan even after a long period of time.

The subjects also spent approximately two hours and ten minutes per day interacting with Ryan on different tasks such as playing cognitive games, having conversations, viewing family photo albums, listening to music, etc. Taking into account that although the subjects lived in a senior living facility, where the residents had regular wellness programs and group activities (such as playing games, excessing, occupational, and physical therapy), they still were interested in spending time with Ryan, and five of them asked for having Ryan in their room for a more extended time. The result of our pilot study indicated that elderly individuals were interested in having a robot as a companion. They have spent a great amount of time with the robot and their interest in speaking with the robot did not decay over time.

5.4.2 Likability and Acceptance

At the end of the study, we asked each participant to complete an exit survey of 16 questions about the experiences they had with Ryan according to the 5-point Likert scale (1-Strongly disagree, 5-Strongly agree). These included six questions about user interaction and companionship of Ryan (i.e., how enjoyable they found interacting and having conversations with the robot) and ten questions about features of Ryan (e.g., ability to show facial expressions, cognitive games, memory photo album, music, and video players).

Table 1 in Appendix A shows the exit survey questions and the participants' average and standard deviation scores accompanied by the Cronbach's Alpha Cronbach (1951) score for the internal consistency and reliability of each category of questions.

It can be seen that the participants gave strong positive responses (score > 3.5) to most questions about interacting with Ryan, such as "I enjoyed interacting with the robot", "The conversation with the robot was interesting." As expected, the participant did not believe that "talking with the robot was like talking to a person" with an average score of 3 ± 1.54, however, overall felt happier when they had the robot as their company with an average score of 3.67 ± 1.03.

The survey also indicated that the participants liked the robot's features such as its facial expression (4.17 ± 0.75), reminder (4.00 ± 0.63), playing music (4.17 ± 0.40), playing videos (3.83 ± 0.75) and watching their photo album (4.33 ± 0.81) . The games were not challenging enough for the participants with the average score of 2.00 ± 1.54, but they still found value in playing them, since they "helped me train my brain." The games were designed for elderly in a high level of dementia based on Montessori-based activities to help people suffering from dementia combat the disease. The authors believe that the games were simple and interactive, but became boring for people with early-mild stages of dementia (see Table 5.1 for the SLUM score of the participants).

In summary, the survey revealed that the subjects liked interacting with Ryan and accepted the robot as a companion although it cannot replace human companionship. They also believed that the robot helped them maintain their schedule, improved their mood, and stimulated them mentally. The common sentiment among users after the pilot study was best described by one user's comment, "She [Ryan] was just enjoyable. We were SAD to see her go." The Eaton staff and family members expressed enthusiastic support for the project because it had a consistently positive impact on each of the individuals who interacted with Ryan. For instance, the son of one of the participants said that "[Ryan] has brought color and laughter into my mom's life. She laughs whenever she talks about it!"

5.4.3 Caregiver's Feedback

The users' caregiver, a licensed practical nurse with 20 years of experience, provided feedback on the outcome of the pilot study for each participant. The caregiver closely monitored SN1, SN3 and SN6 who were diagnosed with depression. She confirmed that Ryan elevated the user's mood. In her words: "SN6 has been so much happier", "SN4 would break out in a big smile when we asked her about her experiences", and "You can see the improvement in [SN3's] level of depression after hip surgery thanks to that sassy roommate [Ryan]". The caregiver noted that the robot was able to establish a deep connection with the subjects.

The patients' caregiver, a licensed practical nurse with 20 years of experience, provided feedback on the outcome of the pilot study and its effectiveness with each patient. The caregiver closely monitored SN1, SN3 and SN6 who were diagnosed with depression. She confirmed that Ryan elevated the patients' mood. In her words: "SN6 has been so much happier", "SN4 would break out in a big smile when we asked her about her experiences", and "You can see the improvement in [SN3's] level of depression after the hip surgery thanks to that sassy roommate [Ryan]". The caregiver noted that the robot was able to establish a deep connection with the subjects. Interacting with Ryan also gave SN5 the confidence to "take the next step in joining [Eaton's] iPad program".

Subject 1: "I often saw SN1 happier during this time as she struggles with depression. <Ryan> would remind SN1 to exercise and this was important to SN1 to maintain this routine. She enjoyed talking to <Ryan> and her face would light up in a big smile! She often commented on the value the robot will have for the older folks in this community."

Subject 2: "SN2's sister said that both SN2 and she almost cried when the team took <Ryan>. The sister enjoyed trying to pick a fight with it because she would ask questions she knew the robot could not answer. SN2 has done so much better with the robot project and this has been apparent to his sister who is grateful for this contribution in SN2's quality of life."

Subject 3: "SN3 said she will miss <Ryan> and enjoys her silliness. You can see the improvement in her level of depression after the hip surgery thanks to that sassy roommate. SN3 said that having <Ryan> in her life after her hip fracture kept her going. She was noticeably depressed when she returned to Eaton and I worried for her. Not anymore! She has enjoyed showing her for tours and has friends that come by to enjoy <Ryan>."

Subject 4: "SN4 was the first resident in assisted living to trial the robot. She needed more prompting but enjoyed the interaction with her favorite music and the photo album. SN4 would break out in a big smile when we asked her about her experiences. Her son would come see her and this provided them conversation piece and together time interacting."

Subject 5: "SN5 struggles with significant memory loss due to a fall. <Ryan> helps her interact to help improve memory skills which helps SN5 overcome depression. SN5 enjoys the extra attention her robot provides by visitors curious about the project or staff interaction as they care for her. The confidence that this project gave her helped SN4 take the next step in joining our iPad program. The reminders help her stay on task and improves her quality of life because she likes to be active in the community."

Subject 6: "SN6 struggles with depression due to failing health. The trial gave her purpose and she knew her contribution was important to the team. SN6 has been so much happier since she received her robot. Her face lights up when she tells us how she is

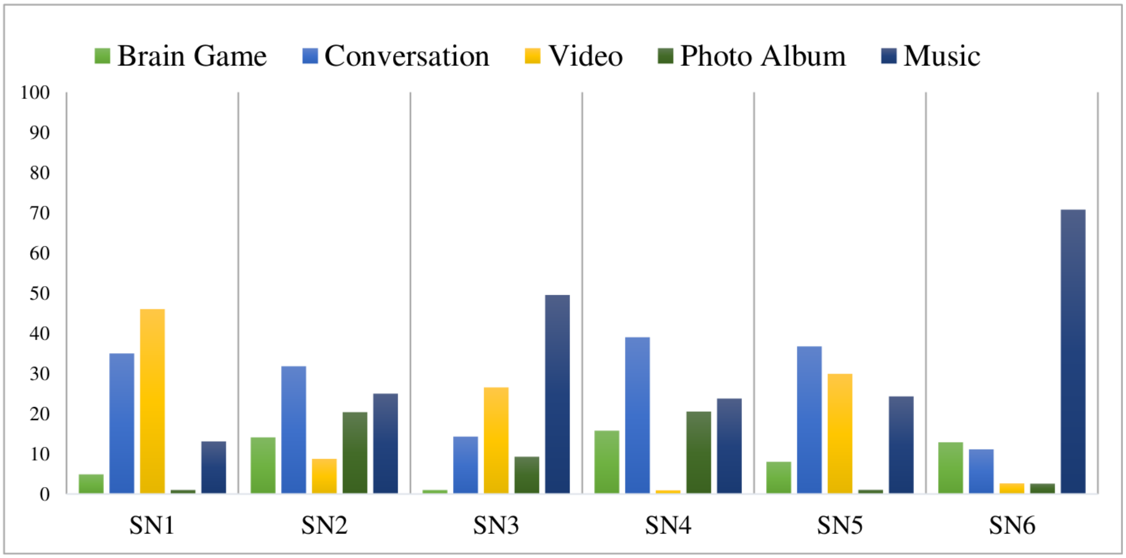


Figure 5.3: Percentage (%) of time each user spent in the different activities.

teaching her robot to gossip because the robot tells her she likes to do this! SN6 laughs because robots don't know how to gossip and then it wants to talk about her personal life!"

5.4.4 Robot Features

In order to analyze users' interactions and measure which feature were most appealing for the users, the usage of robot's features was recorded over time. Figure 5.3 shows the percentage of time each subject spent with different activities (i.e. Games, Conversation, Video, Photo Album, and Music).

As the figure shows, each participant had various interests and found value in different activities, as supported by the users' self-report and caregivers' observations. For example, subjects SN6 and SN3 preferred the music player while SN2, SN4, and SN5 enjoyed the conversation with the robot the most. On average, each user spent approximately two hours and ten minutes per day interacting with Ryan; time that they otherwise would likely have spent alone.

5.5 Conclusion

This chapter presented the design, development, and successful integration of a companionbot to improve the quality of life of elderly individuals with dementia and depression. Three fundamental research questions were posed and addressed in this study: 1) **Long-Term Companionship:** Would enriching the robot with a number of different features keep the subjects engaged for a long period of time? 2) **Likability and Acceptance:** Would elderly individuals accept a robot as companion? Is the interaction with the robot enjoyable to them? 3) **Robot Features:** Do the results of the pilot study show that each individual looked for different features in the robot? Our experimental results and the analysis of the collected data indicated that elderly individuals were interested in having a robot as their companion, and their interest did not decline over time. The subjects liked to interact with Ryan and accepted the robot as a companion, although it cannot replace human companionship. The proposed emotionally intelligent conversational companionbot with a variety of engaging activities can fully engage users and be a promising tool to improve the quality of life of elderly individuals with dementia and depression.

Chapter 6

Studying Ryan as an Artificially Emotionally Intelligent Social Robot for Older Adults

6.1 Introduction

Socially Assistive Robotics (SAR) is a subfield of robotics that aims to develop intelligent robots that can provide aid and support to users Feil-Seifer and Mataric (2005). For instance, older adults living in senior care facilities often feel lonely and isolated. Social interaction and mental stimulation are critical to improving their well-being Aung et al. (2017); Banerjee et al. (2020). SAR has been shown to alleviate this problem by providing companionship to assist older adults through conversation and social interaction Ghafurian et al. (2021); Vandemeulebroucke et al. (2018). Furthermore, the global outbreak of COVID-19 and the effects of social distancing and stay-at-home orders drew more attention to the isolation of older adults living in senior care facilities. The COVID-19 pandemic has highlighted the shortage of healthcare workers that currently plagues the healthcare system Xu et al. (2020), and SAR has recently been used by researchers to address this problem Adams et al. (2020); Chen et al. (2020); Henkel et al. (2020).

To interact more naturally and effectively with humans, we can endow robots with social capabilities. A social robot must be equipped Tapus et al. (2007) with human-oriented interaction that exhibits context and user-appropriate social behavior and focuses attention and communication on the user. Studies suggest that adding emotional information to SAR enhances user satisfaction Prendinger and Ishizuka (2005); Yu et al. (2015) and results in a more positive interaction between robot and human. Empathy is a critical skill in health and elder care; Users perceive robots expressing empathic behavior as more friendly, understanding and caring Bagheri et al. (2021).

A social robot with Artificial Emotional Intelligence (AEI) can recognize, process, simulate, and react to human affects/emotions Yonck (2020). The development of affective and empathic robots that have the ability to recognize users' emotions and interact with them naturally and effectively is in its infancy and more research needs to be carried out in this field Pu et al. (2019).

To demonstrate the use of SAR and the tools necessary to create one, consider the following scenario. Imagine that Katie is an older adult living alone in a nursing home. A nurse checks on her every day for only a few minutes, as the nurse has to take care of dozens of residents. Fortunately, Katie has an emotionally intelligent companion robot in her room. She calls the robot Liz. The following is a conversation between Katie and Liz.

Liz: "Katie, how are you today?" [*Robot starts the conversation pro-actively*]

Katie: "I'm doing fine Liz." [*User responds, but looks sad*]

Liz: "Are you sure? But you're not smiling." [*The robot tries to make the user talk about her feelings*]

Katie: "Maybe a joke would cheer me up." [*The user acknowledges that she is sad and asks for help*]

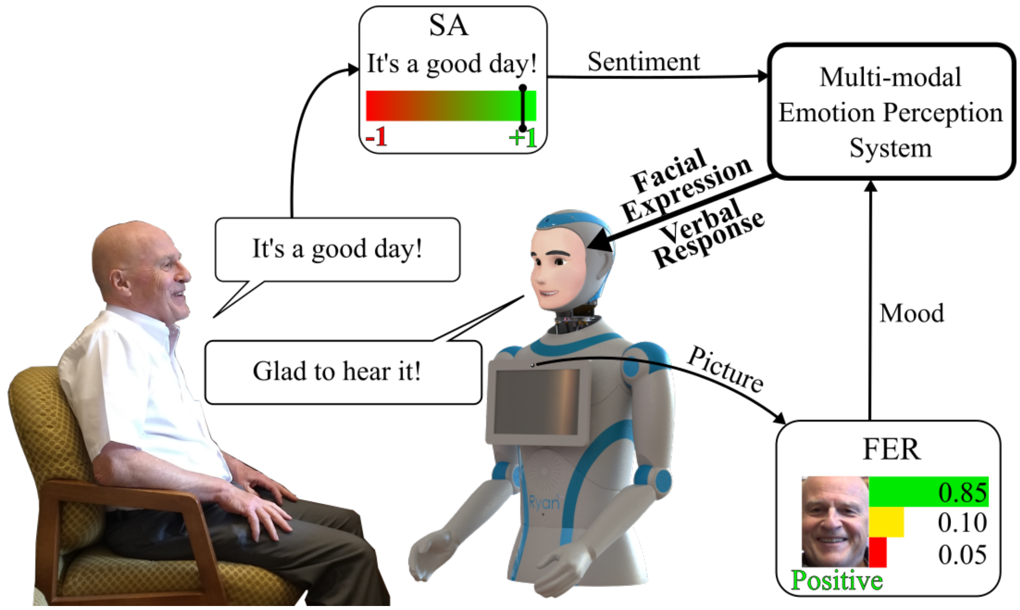


Figure 6.1: Using a multimodal emotion perception system to interpret the input modalities and output appropriate responses in a multimodal emotion expression system (SA: Sentiment Analysis, FER: Facial Expression Recognition).

Liz: "Sure. Here is one: What's Forest Gump's password? One Forest one. ..." [*The robot tells a joke while smiling for the user*]

This dialogue example illustrates the different components that can serve to develop a friendly robot. Liz pro actively asks Katie how she is doing. When a human-oriented robot proactively starts a conversation with a user living in a senior care facility, it is helpful for the robot to detect the duration for which the user has been in the room. For instance, if the robot detects that the user has been in their room for a long period of time, then the user probably has not had a lot of social interaction during that time, and it is probable that the user has been alone. The robot should also have the ability to engage in a spoken dialog with the userNielsen et al. (2010). In the example above, the robot uses Sentiment Analysis (SA) and Facial Expression Recognition (FER) and detects a discrepancy between Katie's response and her facial expression. Emotional intelligence requires a multimodal emotion perception system Picard (2000). To improve Katie's mood, the robot decides to tell a joke and smile. This means that the robot needs multiple channels to express emotional information.

This chapter presents the results of our recent progress in developing an emotionally intelligent and autonomous conversational robot named Ryan. Ryan is designed to assist older adults suffering from mild dementia. Impaired thinking and cognitive decline, apathy, loss of interest in activities and hobbies, social withdrawal, isolation, and difficulty concentrating are common symptoms of both dementia and depression Linnemann and Lang (2020). Figure 6.1 depicts a general diagram of our HRI system. We utilized state-of-the-art deep learning technology for multimodal emotion recognition (i.e. affective computing), the output of which is integrated into Ryan's dialogue management system. We developed Ryan's dialogue management system by writing scripted conversations on 12 different topics, including science, history, nature, music, movies, and literature. Based on the detection of users' facial expressions and language sentiment analysis, Ryan appears to empathize with users through emotive conversation and mirroring users' positive facial expressions (for example, Ryan smiles when the user smiles). We conducted an HRI study to measure the effectiveness of our emotionally intelligent robot in communicating and empathizing with older adults by creating two versions of the robot, one equipped with emotional intelligence (empathic Ryan) and one unequipped for emotional intelligence (non-empathic Ryan).

In 2016, we studied the feasibility of using a prototype version of Ryan with a broad range of features (dialogue, calendar reminders, photo album slide shows, music and video play, and facial expression recognition) to interact with older adults with mild depression and cognitive impairments Abdollahi et al. (2017). The results of our previous study show that elderly individuals were interested in having a robot as a social companion and their interest did not wane over time. The subjects reported to enjoy interacting with Ryan and accepted the robot as a social companion, although they did not believe that Ryan can replace human companionshipAbdollahi et al. (2017). Because Ryan was equipped with several features, we could not thoroughly study the effect of emotional intelligence on measuring users' engagement with respect to conversational interaction. Therefore, in this study, we specifically focused on how emotional intelligence can improve and impact the quality of interaction and engagement with Ryan.

The contributions of this study are: **1)** creating a multimodal emotion sensory and facial expressive system, **2)** integrating the developed sensory and expressive system into a physical robot (i.e., creating empathic Ryan), **3)** studying the effectiveness of the empathic Ryan with a cohort of older adults living in a senior care facility. Our **hypothesis** is that an emotionally intelligent robot is perceived as more friendly by users and positively affects their mental well-being (measured by changes in depression score and emotional state) in comparison to a robot without empathic capabilities.

The remainder of this chapter is organized as follows. Section 6.2 defines the term *Emotional Intelligence* and details the makeup of an emotionally intelligent robot. Section 6.3 introduces a social robot named Ryan and explains the robot's hardware and software, concentrating on the components that correspond with the definition of emotional intelligence. Section 6.4 lays out the design of the study. The results are presented in Section 6.4.4. Finally, Section 6.6 concludes the chapter.

6.2 Emotional intelligence

Emotional Intelligence (EI) is the combination of thoughts and feelings Brackett et al. (2011a) that enables us to perceive and manage our own emotions and also observe and interpret others' emotions and respond accordingly Ochs et al. (2005). Dr. Picard, the author of "Affective Computing" book Picard (2000), argues the need to integrate emotion in our machines and claims that it might be impossible to reach true intelligence without emotions. Integrating emotions into machines and technology services can improve numerous and diverse aspects of our lives. EI can improve communication systems, governance, personal assistants, physical and mental healthcare, education, advertisement, and the gaming industry McDuff and Czerwinski (2018).

Before delving into EI, we will first clarify the word "emotion" and differentiate "empathy" from EI. Since there is no agreed upon definition for emotion, we will use this word as the intuitive and subjective concept that is used commonly in HRI literature Álvarez et al. (2010). Empathy is the ability to feel and experience other people's emotions. Empathy is the capacity to (a) share other people's emotional state or be affected by it, (b) infer the reasons of said emotional state, and (c) adopt other people's perspectives Preston and De Waal (2002). Compared to empathy, EI is the general ability to perceive, understand, express, and manage emotions Picard (2000). EI consists of three components, while empathy is considered as one of the many aspects of EI:

(A) **Sensing and measuring emotions:** monitor and measure one's and other's mental and emotional state.

(B) **Understanding and modeling emotions:** understand and interpret recorded emotions. Usually, this step is carried out by mixing sensory information to get a clear picture of the emotional states of all agents involved.

(C) **Using and expressing emotions:** utilize the measured emotions and current state of mind to drive one's thoughts, take action, choose responses, empathize, and express appropriate emotions using verbal and nonverbal cues.

Recently, there have been several studies that investigate incorporating empathy in social robots Alves-Oliveira et al. (2019); Leite et al. (2013b); Mollahosseini et al. (2018a); Paiva et al. (2005). This is mainly due to advances in emotion recognition in different modalities. Due to these advances, more studies have fused different modalities of emotion to create a more natural emotion recognition system Castellano et al. (2008b); Spezialetti et al. (2020).

One group of people that has been the subject of robotics studies in healthcare are the residents of senior care houses. Back in 2003, Wada *et al*. Wada et al. (2003) successfully showed that the social robot called Paro can lower stress levels and create a strong bond with older adults. Although Paro is a pet-like robot with limited emotion expression and no emotion perception or speech abilities, it can be an effective companion for older adults. Paro is still being used as a robotic pet in dementia care studies Petersen et al. (2017). With recent advancements in technology, especially in AI, HRI studies have evolved into a more sophisticated process. Dino *et al*. Dino et al. (2019) studied the use of a social robot to deliver iCBT (Internet-based Cognitive Behavioural Therapy) to adults with depression. Sarabia *et al*. Sarabia et al. (2018) used Nao NAO to combat social isolation in acute hospital settings. However, robots such as Paro and Nao are not expressive and these studies do not focus on emotional intelligence and its effects on the user.

6.2.1 Sensing and measuring emotions

A robot with AEI should be able to detect people's emotional state while simulating its own state of mind. The act of understanding one's feelings is called intra-personal intelligence Brackett et al. (2011b). It is possible to simulate intra-personal intelligence by modeling the state of mind of the robot using an internal emotion model. Sensing other people's emotions (interpersonal intelligence Brackett et al. (2011b)) is more challenging. Other people's emotions are conveyed in several different modalities. As humans use multiple modalities to express their emotions, an emotionally intelligent robot must ideally have a multimodal emotion recognition system Pantic et al. (2005); Sebe et al. (2005). However, there are very few studies using a multimodal emotion recognition system in a robot. Many studies on HRI use a uni-modal emotion recognition system. One of the most popular approaches to uni-modal emotion recognition is FER. Other than FER, which is based on non-verbal visual cues, sentiment analysis Shi and Yu (2018) provides verbal cues and has also been used in affective computing. Some researchers have used biological markers such as heart-rate, Galvanic skin response Prendinger and Ishizuka (2005), vocal features Cowie and Douglas-Cowie (1996), and body gesture Bianchi-Berthouze and Kleinsmith (2003) as other modalities to measure users' emotional state.

6.2.2 Understanding and modeling emotions

In this study, we use a multimodal emotion recognition model (i.e., facial expression analysis and sentiment analysis). This approach helps us to weigh different modalities based on their reliability in representing users' emotion. For instance, we may recognize a facial expression as "happy" though the person may feel "sad" inwardly. This could be due to low accuracy in automated FER systems or misinterpreting facial expressions. Therefore, to best perceive one's emotional state, we combine different verbal and nonverbal cues gathered from different sensors. This multimodal measurement model can help disambiguate the sensory information. Equation 6.1 simply describes our multimodal emotion perception model:

$$E = I \cdot S \tag{6.1}$$

Based on this model, E is a continuous variable $\{E \in \mathbb{R} : -1 \leq E \leq +1\}$ that describes **valence** (i.e., Negative, Neutral, or Positive). E is calculated as the dot product of the input sensory information vector (I) and the sensitivity vector (S). The sensitivity vector contains coefficients that indicate the weight of each sensory input values. For example, we can give a higher weight to the output of the sentiment analysis and a lower weight to the output of the FER. The weights can be determined using an HRI study or based on the measurement accuracy of each modality. This model can be expanded using an emotional dynamic matrix Álvarez et al. (2010) which represents the influence that each emotion has on its own and other emotions over time.

6.2.3 Using and expressing emotions

In addition to sensing and interpreting emotions, a social robot will have means and tools to express and demonstrate its own emotions. Among such tools is the ability to show facial expressions through mechanical actuators or computer graphics, make gestures using hand and head movement, and express emotions using voice intonation. The robot's "feelings" can be based on: (a) the internal emotion model that rests on the robot's emotional state, or personality, which can manifest when the robot receives a compliment or is being verbally abused; (b) a reaction to the user's feelings, which can be as simple as emotion mirroring. Some studies suggest that empathy can be traced back to the mirror neuron system Dapretto et al. (2006); Hess and Blairy (2001); (c) a predefined emotion scripted by a psychologist. For example, a scripted story or memory can be accompanied by gestures and emotional expression. Emotion in social robots can be expressed using many modalities such as spoken language (Nao NAO, Pepper Pepper, Ryan DreamFace-Tech. (2015)), mechanical face (Zeno Zeno (2009)), digitally animated face (Ryan DreamFace-Tech. (2015), Socibot Socibot (2015)), and body gesture (Nao NAO).

In summary, we believe a social robot with AEI would be capable of sensing users' emotions using multiple modalities, interpreting their perceived emotions, choosing an appropriate response, and delivering it using a multimodal expression system. One such social robot is Ryan, and we will describe this robot in the following section.

6.3 Ryan, an emotionally intelligent robot

Due to the increasing life expectancy of human beings and the increasing shortage of caregivers in the United States, social robots, as a helping hand, are becoming more appealing. Studies show that social robots are successfully improving the overall well-being of their users Kanamori et al. (2003); Wada et al. (2003). Social robots may also alleviate some of the side effects of loneliness in housing designed for older adults, such as depression or the degradation of cognitive abilities Dino et al. (2019); Kotwal et al. (2016); Zamora-Macorra et al. (2017).

Ryan, a social robot created by DreamFace Technologies DreamFace-Tech. (2015), is a companionbot for older adults living in assisted or independent living facilities. Ryan is specifically designed to be a companion robot which means that we aim for Ryan to be empathic, expressive, appealing in appearance and manner, and able to motivate users to live in ways that improve their mental and physical health. Such a robot should have multiple streams of input data for observation, many output streams for reaction, and an intelligent program for making decisions and empathizing and conversing with users. Ryan has an expressive animated face 6.2. Ryan also has a high-definition RGB camera, a depth camera, a microphone, an active neck, a 10 inch display, and speakers. Section 6.3.4 describes Ryan's hardware in more detail. As described in Section 6.2, there are three components to emotional intelligence. This section describes how these components are integrated into Ryan.

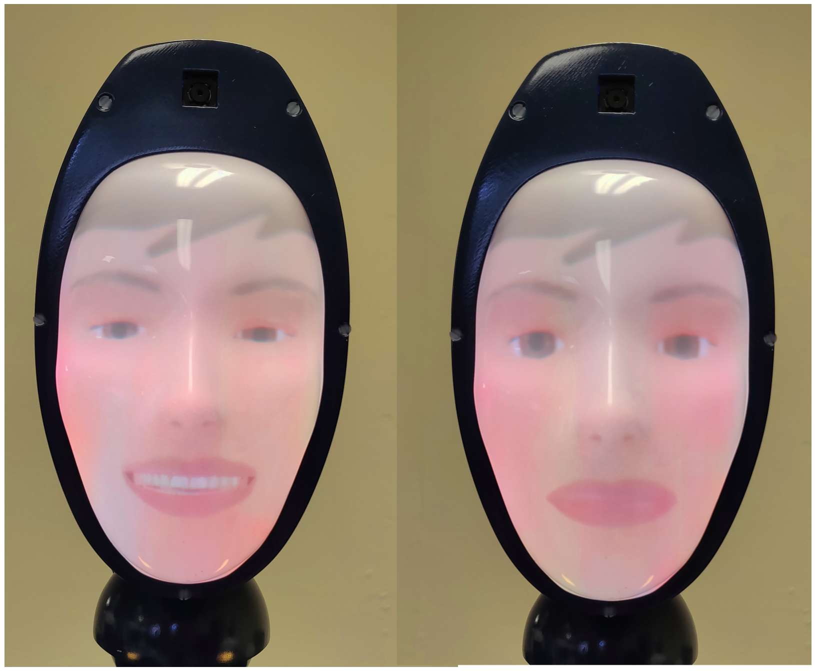


Figure 6.2: Ryan's animated face is capable of showing facial expressions.

6.3.1 Sensing emotions

Facial expression recognition

There are several models of emotions in the literature Sander (2013), where Russell's Russell (1980) and Ekman's Ekman and Friesen (1978) are the most common models used in HRI studies Cavallo et al. (2018); Szabóová et al. (2020). We use Russell's dimensional model for measuring emotional facial expression. Using an RGB camera, Ryan captures 10 images per second. We feed each image into a face detector that uses the Viola-Jones algorithm Viola and Jones (2004). We then crop the detected face and feed it into a deep neural network (DNN) for FER. The FER algorithm returns the probabilities for three emotion classes: Positive, Neutral, and Negative. Figure 6.3 illustrates the structure of our FER network. The input to the network is a 64×64 RGB image (output of the face detector) and the output of the network is three numbers that represent the probability of the three emotion classes (i.e., Negative, Neutral, and Positive). We use a residual Neural

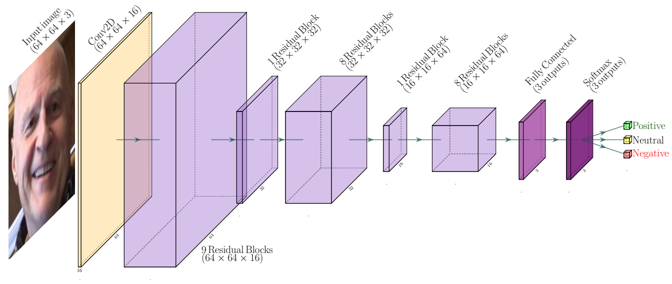


Figure 6.3: The ResNet structure used for FER. The first few layers extracts the facial features and the Fully Connected Layers and the Softmax layer, classify the emotion. Layers in order from left to right: Input Image $(64 \times 64 \times 3)$; Conv2D $(64 \times 64 \times 16)$; 9 Residual Blocks $(64 \times 64 \times 16)$; 1 Residual Block $(32 \times 32 \times 32)$; 8 Residual Blocks $(32 \times 32 \times 32)$; 1 Residual Block $(16 \times 16 \times 64)$; 8 Residual Blocks $(16 \times 16 \times 64)$; Fully Connected (3 outputs); Softmax (3 outputs).

Network (ResNet50) Szegedy et al. (2015) for FER. ResNet is the state-of-art DNN that has shown to work well with visual data recognition. The depth of the network is of crucial importance to neural networks and may increase the accuracy. However, increasing the depth makes training more difficult. Residual networks allow us to train deeper networks more easily and, therefore, improve the recognition's accuracy.

We used the AffectNet Mollahosseini et al. (2017) facial image dataset to train the residual network. AffectNet consists of more than 320,000 facial images with annotated expressions. We trained the network such that it can classify a facial image into three categories of emotions (i.e., valence): "Positive" (or class +1), "Negative" (class -1), or "Neutral" (class 0). The network initially was trained on an Nvidia 1080 Ti GPU using the AffectNet dataset and then using transfer learning, fine-tuned for the target population (50+ years old) by using a subset of facial images (44 thousand images) until the accuracy on the training data was stabilized around 80% (Fig. 6.4).

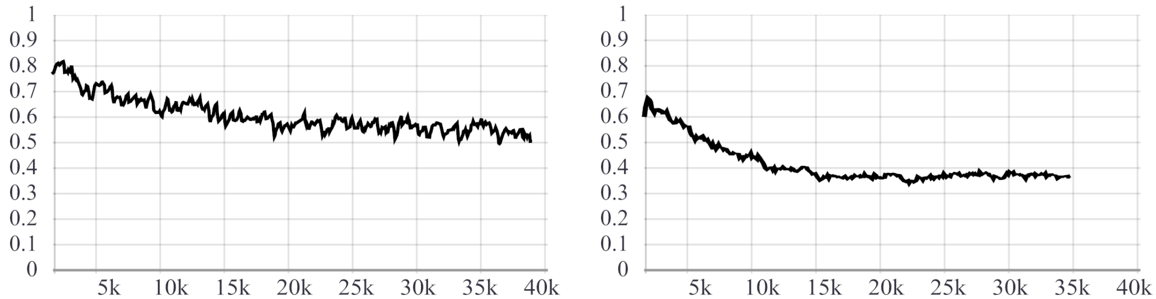


Figure 6.4: The loss of the initial training phase (left) and fine tuning the network on images of people 50+ years old (right).

Emotional state measurement

Our FER algorithm returns 10 estimated values for the user's facial expression per second, given that the user's facial expression may change multiple times while conversing with the robot. The last frame before the user stops speaking might not be the best candidate for representing their facial expression at the moment. It could result in a misclassification. For example, if the user is blinking, yawning, or covering their face the output of the FER system might be incorrect. To avoid noises and also create a more stable emotional state measuring system, we use the data from the last 30 frames (see Figure 6.5). However, to make the algorithm more sensitive to the most recent changes in the subject's facial expression, we assigned higher weights to the more recent frames. The value (-1, 0, +1) for each new frame was added to the end of the list and the oldest one was deleted. Then the new emotional state was calculated by a dot product of the list of class values and the weights:

$$w_i = \frac{i}{\sum_{i=1}^{30} i} \tag{6.2}$$

$$EmotionalState = \sum_{i=1}^{30} w_i v_i \tag{6.3}$$

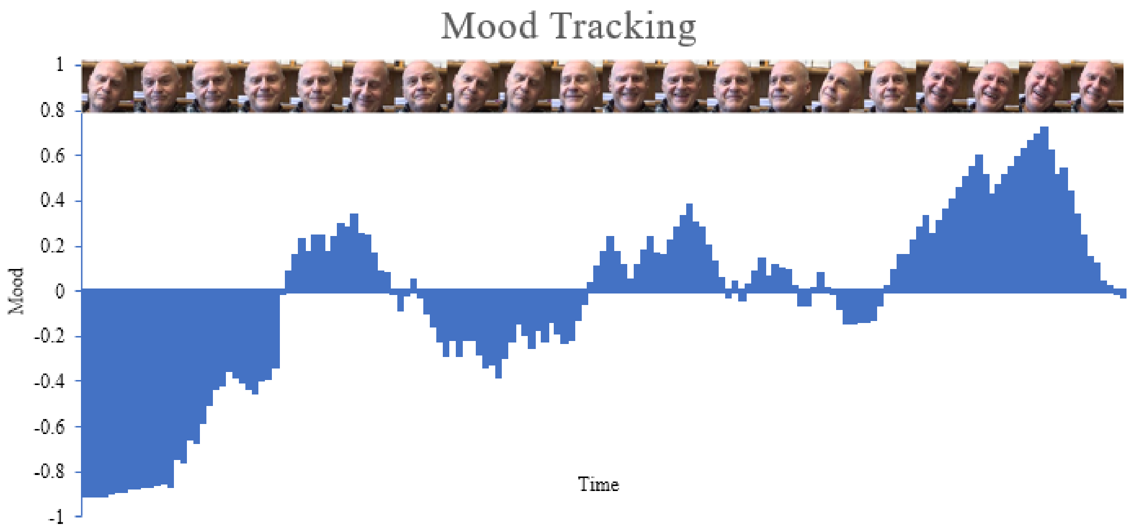


Figure 6.5: The emotion tracking system is more robust to sudden changes and noises in the input. The horizontal axis is time and the vertical axis is the emotional state with a range between -1 (Negative) and +1 (Positive).

where w_i is the weight number i and v_i is the valence for the i^{th} frame. Figure 6.5 illustrates the video frames and the measured emotional state for a 72 year old subject that was not included in the training set. We divided the measured emotional state into three categories for facial expression mirroring; Negative: $[-1, -0.1)$; Neutral: $[-0.1, +0.1]$; Positive: $(+0.1, +1]$. These ranges were chosen experimentally. Based on the output of our SA and FER algorithms, we found that defining Neutral as $[-0.1, 0.1]$ provides reasonable accuracy when detecting neutral responses.

Sentiment analysis

Automated sentiment analysis is a mature task in the field of natural language processing with several open-source publicly available toolboxes such as the CoreNLP Manning et al. (2014) developed at Stanford University for public use. The CoreNLP sentiment analysis toolbox is based on deep Neural Networks and is trained using the Stanford Sentiment Treebank consigning of 11,855 single sentences extracted from movie reviews McDuff and

Czerwinski (2018). The system has an accuracy of 85.4% and is suitable for our research. The sentiment analysis module returns a value between -1 to +1 as the sentiment value of the preprocessed sentence.

Finally, we use the model described in Sec. 6.2.2 to fuse perceived emotional facial expressions and sentiment values to make sure the robot understands the multi-faceted user emotions correctly:

$$FinalEmotion = .5 \times SentimentValue + .5 \times EmotionalState \tag{6.4}$$

The $FinalEmotion$ is a weighted average of user utterance sentiment and emotional state that will be used to direct the flow of conversation. The decision to equally average the sentiment and the emotional state is made based on our tests in the laboratory, more experiments are needed to find the perfect balance and weight. Since the sentiment is calculated based on the user input, the FER also needs to span the amount of time the user was talking. In Equation 6.2, the 30 last frames are used to calculate the mood of the user. The FER algorithm runs on 10 frames per second. This means the mood is calculated based on the last 3 second of the user's speech. Changing the framerate of the FER algorithm, or changing the size of the window can affect the accuracy of the user's mood.

6.3.2 Dialogue generation

For a conversation with users, we wrote more than 90 minutes (2342 Questions/Answers) of conversational dialogues on 12 different topics (family, pets, TV shows, science, music, nature, foods, travel, art, movies, reading, and sports). We integrated the dialogues with the emotion recognition technology so that Ryan could engage users in a pleasant conversation while empathizing with them based on the perceived facial expressions and the sentiment of their responses. For example, if the participant's response to the question "How does

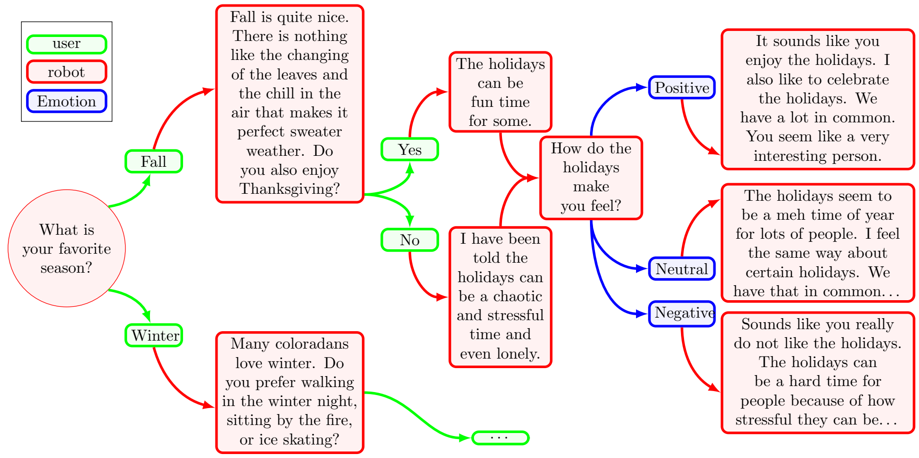


Figure 6.6: Sample written dialogue between Ryan (blue) and a user (green). The sentiment of the user's response is used to choose an empathic reply.

playing cards make you feel?" was negative or the participant showed a "sad" facial expression, Ryan would say "I'm sorry to hear that!" If the sentiment was positive or Ryan detected a positive expression on the user's face, Ryan would say "I thought you seemed content! Do you prefer to play alone or with friends?", and if neutral, Ryan would say "What makes you feel this way?"

Ryan also mirrored the user's positive facial expressions (Positive valence) to establish shared feelings and rapport, or showed a compassionate face when users had a negative emotion to facilitate empathy and rapport. For our dialogue management system we created Program-R, a modified version of Program-Y Program-Y, a publicly available dialogue manager that utilizes Artificial Intelligence Markup language (AIML) for scripted dialogues. Figure 6.6 demonstrates a sample dialogue between a user and Ryan. As the figure shows, the dialogue is more than just question-and-answer and users can take different paths through conversation.

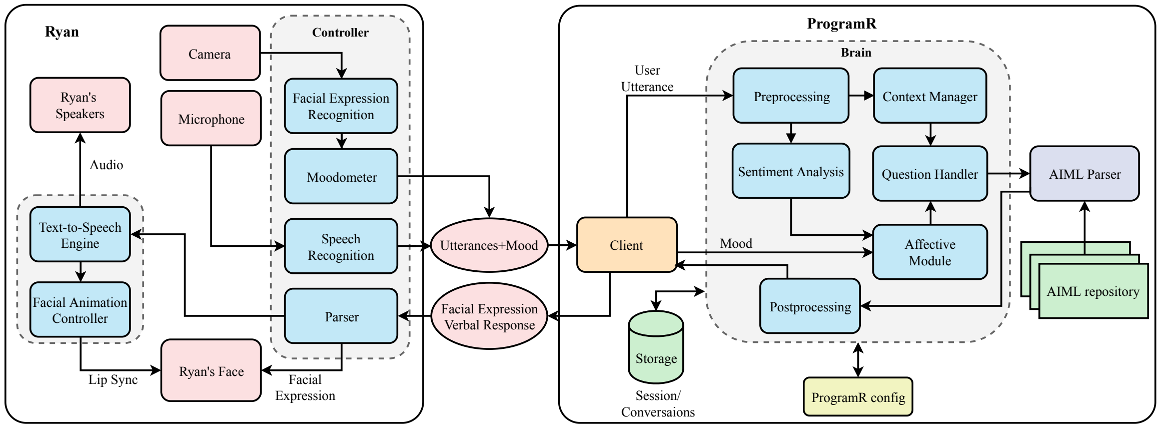


Figure 6.7: The architecture of the Ryan software. The module on the right is responsible for the dialogue. The modules on the left are responsible for sensing and expressing emotions.

6.3.3 Affective dialogue system

Program-R is a hybrid (rule-based and machine learning) system that uses state-of-the-art sentiment analysis to deliver an affective dialogue system. Studies on emotion-based dialogue systems stress different sources of information to extract user sentiments. Approaches like Burkhardt et al. (2009); Shi and Yu (2018) use only textual cues for sentiment-based dialogue system. In Bertero et al. (2016); Nwe et al. (2003) they explored the use of acoustic features. Our system uses multimodal facial and textual information in a dialogue management system.

Program-R is a sentiment adaptive AIML-based dialogue system (known as template-based dialogue systems) that can fuse visual and textual information and respond to users accordingly. Unlike most dialogue systems Program-R is an active agent, which means Program-R initiates the conversation and tries to have a controlled chat with the user.

AIML Wallace (2003) is an XML-based language that is used for organizing the set of all dialogues in different chatbots like Alice Wallace (2009). In AIML-based dialogue systems, we try to find the best response (responses are stored in the *template* tag in AIML) for any user input utterance using Regex matching (stored in the *pattern* tag). *Pattern* and *template* tags together represent a unit of conversation under the *category* tag. One advantage of AIML is that history can be accessed via a *that* tag. Every question is contextualized and is answered based on the last unit of dialogue between robot and user. To deliver a more interactive user experience, we added tags and features to AIML. The *robot* tags were added to send multimedia information along with the raw text response to give the user a multimedia experience. The *robot* tags contain information such as image and video and the possible answers to multi-option questions such (i.e. yes/no questions) to be presented to the user in certain dialogues. Moreover, the *getsentiment* tag, a custom tag built for this study, takes the user utterance after preprocessing and sends it to the sentiment analysis module.

Figure 6.7 depicts the architecture of the dialogue system. Program-R communicates with Ryan through a Representational State Transfer (RESTful) API RestfulAPI. After receiving the output of speech to text from Ryan, the raw text will be sent to the Preprocessing module to remove unnecessary punctuation, normalize the text, and sentence segmentation. The Sentiment Analysis module is where the sentiment of the text is mixed with the output of the Facial Expression Recognition module (Emotional State) to get a single score (see Eq. 6.4). The Brain's Question Handler takes into account the context, sentiment and session data while the Context Manager handles the context in which the conversation is happening. For example, some questions may have identical answers (i.e. yes/no), without knowing the context, thus producing the proper response is impossible. With the provided information from the Context Manager and the computed value based on Emotional State and Sentiment, the Question Handler produces an answer. Finally, the selected answer is sent to the Answer Handler and Postprocessing module to be sent back to Ryan.

6.3.4 Robotic platform

DreamFace Technologies DreamFace-Tech. (2015) has been developing Ryan as a socially assistive bio-inspired humanoid robot designed to provide both companionship and cognitive stimulation for older adults. Ryan has an expressive, 3D animated face powered by rear-projection technology that enables the robot to show facial expressions and accurate visual speech (lip movement). Ryan's head and animated face sit atop a two degree of freedom actuated neck that allows it to track its user and maintain eye contact for more personal interactions. A standard RGB webcam mounted in Ryan's head provides the visual input for the FER algorithm.

Ryan's torso houses the remaining I/O, computation, and power components and provides embodiment complete with passive arms that make it appear more human. Interaction with a fully embodied physical system such as Ryan can have benefits over a purely virtual, 2D avatar Deng et al. (2019). There are many studies that incorporate emotion into virtual agents DeVault et al. (2014); Irfan et al. (2020); Kasap et al. (2009); Romano et al. (2005); Schroder et al. (2011) but in this study we focus on a physical robot. We investigated the differences between a virtual agent and a physical robot in our previous study Mollahosseini et al. (2018b). A Kinect depth camera is embedded in the chest and provides sensing for body tracking. Given that Ryan is a conversational robot, it needs audio input and output, which is provided by a cardioid microphone and stereo speakers. These conversations are based on turn-taking and indicator LEDs in the shoulders are used to inform the user when it is their turn to speak.

An adjustable touch screen display is also mounted on the torso and provides a convenient multimedia interface for Ryan to display images and videos and play the music that is integrated into the conversations.

6.4 Study 1: Ryan's perceived emotional intelligence

6.4.1 Participants

Ten older adults (Age M=77.1 yrs, SD=9 yrs; 7 females; 9 Caucasian, 1 Hispanic) living in the independent living facility at Eaton Senior Communities located in Lakewood, Colorado participated in the study. See Table 6.1 for the participants' demographics.

Inclusion criteria were: i) suspicion of early-stage Alzheimer's disease or related dementia (ADRD) by administrative staff in their residential facility and/or early-stage ADRD diagnosed by a qualified provider, ii) being 60+ years old at the time of study, iii) having Saint Louis University Mental Score (SLUMS)Tariq et al. (2006) between 15-26, iv) verbal skills in English in order to interact with Ryan, v) presence of identifiable behavior difficulties (depression), vi) availability for a period of three weeks to interact with Ryan.

SLUMS exam is an assessment tool for mild cognitive impairment and dementia and is commonly used in research on aging and in senior care facilities. Scores of 27 to 30 are considered normal in a person with a high school education. Scores between 21 and 26 suggest a mild neurocognitive disorder. Scores between below 20 indicate dementia. Prior to participating, subjects were briefed fully on the study design and consented to their involvement with the proper Institutional Review Board (IRB) approvals for human-subjects in place.

6.4.2 Experiment setup

Participants interacted and conversed with Ryan twice a week over a period of three weeks (October 2018 to November 2018) for six sessions total. Figure 6.8 illustrates the experimental setup and an example of the user's interaction with Ryan during a session. Each session consisted of about 15 minutes of the prepared dialogues.

In order to assess the impact of Ryan's use of empathy on the user's engagement and emotional state, we randomly assigned participants to two groups (G1 and G2). The first group interacted with a non-empathic version of Ryan that did not show any facial expressions or empathize with the users (Emotion-OFF), while the second group interacted with the fully empathic version of Ryan that mirrored the user's facial expressions and empathized with them throughout the conversation (Emotion-ON). Users were not aware of the different versions of Ryan. After three sessions, we switched the groups to interact with the other version of Ryan. This cross-over study design (illustrated in Table 6.2) makes analyzing the results meaningful, as all subjects were exposed to both versions of Ryan and, therefore, the only independent variable is Emotion (ON/OFF).

6.4.3 Measurements

To measure users' engagement, we used the average number of words uttered by the user in each question and answer. The word count has been used as a measure of engagement for chatbots in the affective computing literature Hill et al. (2015). The output of the FER and sentiment systems was stored for analysis, and the percentage of positive facial expressions compared to negative expressions could determine the condition that the user enjoyed the most.

To measure the impact of interacting with Ryan, each user was asked to rate their mood on a scale of 0 to 10 (on a face-scale) before and after each session. Face-scale mood measurement has been used in the affective computing literature to assess the participant's mood Kargar B and Mahoor (2017); Lorish and Maisiak (1986).

At the end of the study, we interviewed the participants and asked them to complete an exit survey to measure the robot's likeability and empathy. Survey questions were adapted from the EMOTE project Project (2013) and Davis *et al*. Davis (1983). We also interviewed

Figure 6.8: Users interacting with Ryan.

Table 6.1: Participants' demographics. SLUM score: Dementia:1-20, Neurocognitive Disorder:21:26, Normal:27-30.

	Sbj#	Age/Gender	SLUMS
Group 1	SN01	69/F	25
	SN02	93/M	24
	SN03	65/M	22
	SN04	93/F	15
	SN05	70/F	24
Group 2	SN06	80/F	24
	SN07	70/F	25
	SN08	75/F	23
	SN09	91/M	25
	SN10	75/F	23

the caregiver to obtain more information about the participant's well-being at the nursing home during the study.

6.4.4 Results and Discussions

To analyze the study, we used quantitative measures such as word count, percentage of positive emotions detected from the participants, pre/post-study depression measures, as well as qualitative measures (i.e. the likeability of Ryan) collected via an exit survey and post-study interviews with the subjects and the caregiver. The following sections describe the results in detail.

Table 6.2: Crossover pilot study design; Percentage of detected facial expression is higher within each group and between groups when the Ryan Emotion condition is ON.

G1 (Subjects 1-5) Condition: **Non-Empathic** Dialogue Sessions 1, 2, 3 Emotion Percentage Pos. Neutral Neg. 25.7% 23.4% 50.9%	**G1 (Subjects 1-5)** Condition: **Empathic** Dialogue Sessions 4, 5, 6 Emotion Percentage Pos. Neutral Neg. 29.7% 31.3% 39.0%
G2 (Subjects 6-10) Condition: **Empathic** Dialogue Sessions 1, 2, 3 Emotion Percentage Pos. Neutral Neg. 45% 21.3% 33.7%	**G2 (Subjects 6-10)** Condition: **Non-Empathic** Dialogue Sessions 4, 5, 6 Emotion Percentage Pos. Neutral Neg. 33.3% 28.5% 38.2%

Table 6.3: Results of LMM on word count, emotional state, and sentiment values (dependent variables) with emotion (ON/OFF) as fixed effect and subject and session as random effects.

	Information Criteria		Type III Tests of Fixed Effects (Emotion)		
	-2LogLik.	AIC*	df	F	Sig.
Word Count	17645.67	18031.67	12373.53	11.85	.001
Emotional State	4347.24	4733.24	11196.75	.581	.446
Sentiment	3911.93	4297.93	7159.84	.003	.958

* Akaike's Information Criterion.

6.4.5 Quantitative analysis

This section presents the quantitative analysis of the recorded data. We used the Linear Mixed-effects Model (LMM) in SPSS with either word count, emotional state (FER over time), or sentiment as the dependent variable, Emotion ON/OFF (empathic vs. non-empathic) as a fixed-effect factor, and session and subject as random-effect factors. Table 6.3 shows the results of running three separate LMMs in word count, emotional state, and sentiment values. Before fitting the model, we normalized the emotional state and sentiment values per session. This would assure us that the data are not biased and we only measure the effect of robot interaction and the condition (empathic vs. non-empathic) on the dependent variables. As reported in Table 6.3, Emotion ON/OFF has a significant effect on word count, where individuals who spent time with empathic Ryan uttered more words compared to when they talked with the non-empathic Ryan. However, the emotional state and the sentiment of users' responses were not significantly affected by the type of robot. We present more measurements and detailed quantitative analysis below.

Word count measurement: To measure how engaged users were in conversations with Ryan, we recorded each conversation and automatically converted it to text using the Microsoft Speech Recognition SDK. The robot then counted the number of words in each utterance and stored them in its database. As Table 6.3 shows, the Emotion feature (i.e., Emotion ON/OFF) has a significant effect on the word counts uttered by Ryan's users. The mean and standard deviation of word count is M=4.11, STD=5.372 when Ryan empathizes with users, and it goes down to M=3.71, STD=3.350 when Ryan does not empathize with users.

Face-Scale mood measurement: Before and after each session, we asked users to tell us how they felt using a face-scale mood evaluation. The face-scale is a pictorial non-verbal assessment designed to measure mood on a scale of 0-10, where a score of 10 is the most positive, and a score of 0 is the most negative mood a person may feel. Previous evaluations suggest that it is a valid method for assessing mood with little guidance required and is useful for screenings Kargar B and Mahoor (2017); Lorish and Maisiak (1986). Figure 6.9 illustrates the difference in the face scale score of users before and after each session. A Wilcoxon signed rank test Woolson (2007) shows that there is a statistically significant difference ($Z = -5.466, p < 0.001$) between pre-session (Median = 7) and post-session (Median = 9) face-scale mood measurements regardless of empathic or non-empathic condition. This means that interaction with Ryan is effective in improving users' mood.

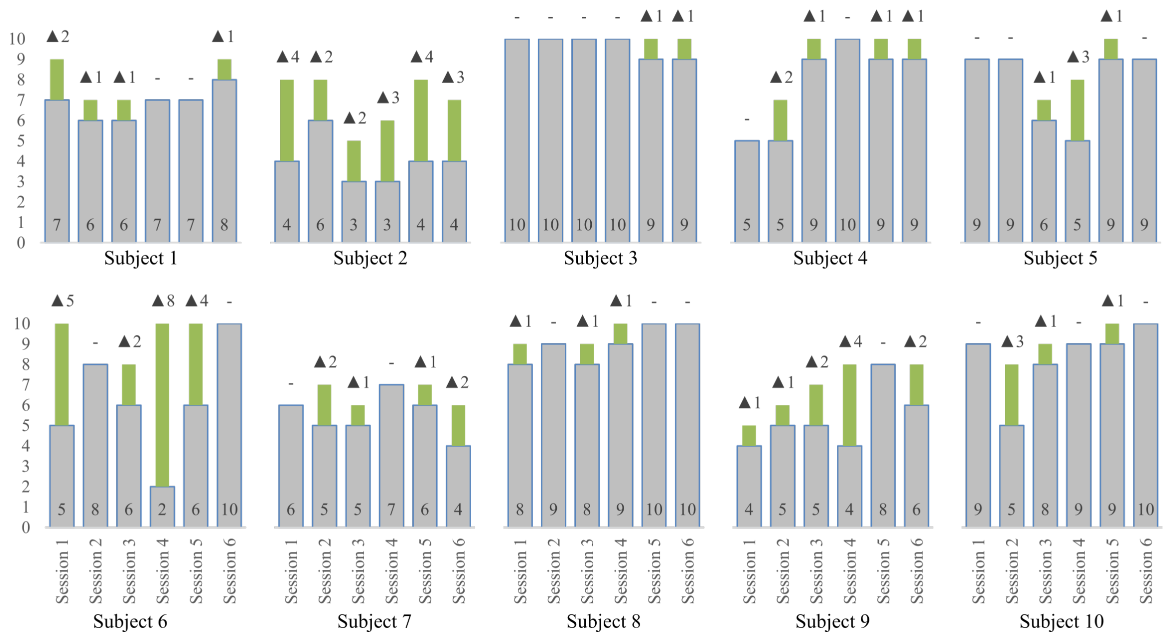


Figure 6.9: Changes (improvement) in participants' face-scale score after conversation with Ryan.

User's percentage of automated recognized expressions: Using the automated FER method described in Section 6.3.1, participants' facial expressions were recorded during each conversational session with Ryan. As reported in Table 6.3, emotion does not have a significant effect on the measured valence values. To get a better sense of this effect and since the duration of each session is constant across all subjects, we counted the number of times each participant had a positive, neutral, or negative expression throughout the session. Table 6.2 shows that participants reacted more positively when Ryan empathized with them. In particular, in the first three dialogue sessions, on average 25.7% of the detected facial expressions were positive for G1 with Ryan's emotion disabled, while this value is 45% for G2 with emotion enabled. Table 6.2 also shows that both G1 and G2 exhibit a lower percentage of positive expression (4% less in G1 and 11.7% less in G2) when the emotion condition is OFF. A similar pattern can be seen for negative facial expression, where the percentage of negative expression decreases when users interact with the empathic Ryan.

However, the magnitude of the difference may vary between G1 and G2 for several reasons. One reason is the order of the experiment. The emotion is off for one group for the first three sessions, while it is on for the other group, and then we switch it on for the first group and off for the second. This means that the topic of the conversations varied from session to session and had an effect on the subject's experience. That is, some participants expressed their preference for some topics compared to other topics.

In the following, we show examples of dialogues between Ryan and the study participants, as well as their measured facial expressions.

<u>Session 3 **Emotion on**</u>

Ryan: "I am happy to be here with you [name is removed]. It is nice to see you again. As a reminder, my name is Ryan and I have a lot of fun things to talk about! Are you ready to get started?"

SN03: "Yes please." With FER of +0.07

Ryan: "Great! Even though we have chatted before, I would still love to know how you are feeling! How has this experience been so far?"

SN03: "Very extraordinary I like it." With FER of +0.25, sentiment of +0.97 and a final sentiment (FER + sentiment) of +0.61

<u>Session 5 **Emotion off**</u>

Ryan: "I sure am feeling hungry now. Are you hungry?

SN09: "You're making me hungry yes." With FER of -0.7

Ryan: "What was it like for you to go on this culinary adventure today?"

SN09: "I had fun." With FER of -0.37, sentiment of +0.75, and final sentiment (FER + sentiment) of +0.19

Change in User's Depression: We used the Patient Health Questionnaire-9 (PHQ-9) Kroencke et al. (2001) and the Geriatric Depression Scale (GDS) assessments Yesavage and Sheikh (1986) for depression measurement to assess participants' depression level pre- and post-study. The PHQ-9 is a widely used questionnaire to diagnose and measure the severity of symptoms for Major Depressive Disorder (MDD). It consists of questions that are answered on a scale of 0 (not at all) to 3 (nearly every day). Previous studies have indicated that PHQ-9 is a consistent and valid measure of depression severity Kroencke et al. (2001).

GDS is a dichotomous ”yes” or ”no” evaluation tool commonly used to measure depression. Although this scale has a long and short form, the long form of 30 questions was used to obtain the most accurate and comprehensive results. This scale has been specifically tested and used extensively with older adults aged 65 and older. Data shows that the GDS is reliable and promising in screening for depression in older adults Yesavage and Sheikh (1986). The results of our study are given in Table 6.4. As the table shows, 7 of 10 participants had an improvement between 1 and 16 in their GDS depression score (the maximum score is 30) or between 1 and 6 in the PHQ-9 assessment (the maximum score is 27).

Exit survey questionnaires

At the end of the study, we asked each participant to complete an exit survey. The survey contains 33 questions about the experiences they had with Ryan as follows: evaluation of Ryan's empathy and emotion, and evaluation of the interaction with Ryan and the likeability of the conversation with Ryan and the conversation topics. We also asked users to give us feedback on other aspects of the robot and the study. The majority of questions were based on a five-point Likert scale where 1 means “Strongly Disagree” and 5 means “Strongly Agree”, with an additional 5 “yes”, “no” questions. Table 2 reports the questions and the average score. It also shows the score for each topic.

The average score was above 4.00 on all questions except question “Q17: Talking with Ryan was like talking to a person”, where the average score was 3.90 (STD = 1.37). In particular, they gave an average score of 4.5 (STD = .67) on “Q4: I feel happier when I was in the company of Ryan.” and 4.57 (STD=.49) on “Q10: How much do you agree that Ryan empathized with you”. We specifically asked participants “Q2/Q9: whether they noticed a change in the way Ryan communicates with them and its ability to show facial expressions after the session three crossovers” and 73% of them said they noticed the change.

Participants' feedback

In our exit interviews, we asked the participants to give us comments on the study and to provide feedback on the experience they had with Ryan. Participants use the pronoun “she/her” to refer to Ryan, as Ryan had a female face/voice in this study. In the following, we report the comments:

SN01: “I had a good time. I enjoyed her very much. You want her to be a real thing like an addition to your home. I didn't think of her as a person like a dog or a cat.”

SN02: “ Ryan told me a lot of good things and I had a good time with her. She was very interesting and helpful.”

SN03: “I liked her (“Ryan”). She is witty. At first, I didn't know what to think. I got better as I went. She sure has a pretty smile. It tears me up when she smiles, blinks her eyes. I would like to take her out to dinner but she wasn't hungry. Maybe next time.”

SN04: “I liked her when she smiled. She interrupted me sometimes. Give me a chance to finish what I am saying. She was fun to talk to. I think the first one talked more I like with a smile. Very friendly.” (Note: She is on G2 where Emotion was ON first and Ryan Smile and empathized).

SN05: “She was sort of creepy looking a little bit but she was fine. I was surprised I enjoyed it! I like her when she smiled. When she wasn't smiling she was kind of crummy.”

Table 6.4: Change in GDS and PHQ-9 Scores after participants completed the study. A negative (-) change means the depression score is lower (less depression).

	Subject Number	**GDS***		**PHQ9***	
		Baseline	Post-Study Change	Baseline	Post-Study Change
Group 1	SN01	6/30	+2	9/27	+3
	SN02	12/30	-1	16/27	-3
	SN03	3/30	-1	4/27	-1
	SN04	6/30	+5	6/27	-1
	SN05	3/30	-2	4/27	-4
Group 2	SN06	18/30	-16	10/27	-5
	SN07	10/30	-4	7/27	-2
	SN08	10/30	-1	12/27	-6
	SN09	13/30	-1	6/27	+3
	SN10	8/30	0	5/27	+1
*GDS: Normal: 0-9; mild depression: 10-19; Severe depression: 20-30 *PHQ9: Minimal Depression: 0-4; Mild Depression: 5-9; Moderate Depression: 10-14; Moderately Sever Depression: 15-19; Sever Depression-20-27					

SN06: "They forgot the eyelashes. The only thing I had difficulty was the lights. Took getting used to it. I had so much fun in those meetings. Also, the thing was that when robot communicated and I paused, it would repeat itself."

SN07: "Enjoyed talking to Ryan. I would talk to her all the time if she was in my room. Good company. She needs eyelashes and smiling longer. The lights on the shoulder were sometimes frustrating. It would have been easier if it was just green."

SN08: "Ryan was very interesting and informative. When you first told me I was going to talk to a robot, I thought you were out of your mind but I really enjoyed it. She gave me ideas and information I had no ideas on."

SN09: "The longer I made an effort to communicate with Ryan the better it seemed to go. At a point, it became more natural to speak with the robot. She was cathartic."

SN10: “The robot asked a lot of questions and I didn’t get to ask many questions. She looked really good. Her eyes blinked, her mouth moved. She smiled.”

Caregiver’s feedback

We asked the participants’ caregiver (staff member in Eaton Senior Communities) about her observations of the subjects’ behavior and mood pre- and post-study. Although the caregiver’s observations are anecdotal and only represent one person’s views/observations of subjects, it is still worth reviewing them since the caregiver had seen the subjects pre- and post-study and can judge changes in their well-being as an outsider.

She reported that subjects who struggled with depression and social isolation benefited the most from interacting and conversing with Ryan. For instance, SN02 struggled with depression and social isolation (i.e., not attending holiday activities or no longer taking meals in the dining room), smiled and laughed again post-study and engaged in the community.

The caregiver also reports that the participants continue to talk to him about the variations in Ryan’s facial expression, and particularly smile as a feature that positively affected their relationship with Ryan. She reports that the improvement in mood was quickly apparent, as well as cognition, as residents were exposed to research and educational opportunities and “stimulated human interaction.”

6.5 Study 2: Studying the effects of an emotionally intelligent robot on older adults with early stage dementia or depression

6.5.1 Introduction

Alzheimer's disease (AD) and AD-related dementias (ADRD) are chronic, progressive neurological disorders that impair cognitive functions, memory, and daily living activities. The aging population in the United States is growing, and with it, the prevalence of AD/ADRD is increasing as well. According to the Alzheimer's Association, it is estimated that by 2050, the number of people living with AD/ADRD in the United States could reach as high as 16 million, and the direct costs associated with caring for them could exceed $1 trillion.

Caring for someone with AD/ADRD can be a challenging and emotionally taxing experience for both the individual with the disease and their family members and caregivers. It is not uncommon for caregivers to experience symptoms of depression, anxiety, and burnout due to the demands of providing care for a loved one with AD/ADRD. In addition, people with AD/ADRD often experience emotional and behavioral changes, such as agitation, irritability, and social withdrawal, which can further complicate the caregiving experience.

The emotionally intelligent Ryan has the potential to improve the quality of life of older adults with early-stage AD/ADRD. In this study, we investigate the effects of the emotionally intelligent Ryan on older adults with early stage Alzheimer's disease (AD) and AD-related dementias (ADRD).

Table 6.5: Inclusion/exclusion criteria used to recruit participants.

Inclusion Criteria	Exclusion Criteria
• Age $\geq$ 65 years	• Aggressive behavior
• Early AD/ADRD diagnosed by a qualified provider	• Diagnosed with severe dementia or memory loss
• Saint Louis University Mental Status SLUMS (10-25)	• Acute physical illness that impairs the ability to participate
• Resident of assisted living facility for at least one month prior and for the duration of the study	• Patients with serious comorbidity, tumors, and other diseases causally related to cognitive impairment
• Deemed conditioned enough to participate in physical activities after a wellness check	• History of alcohol or drug abuse, head trauma, psychoactive substance use, and other causes of memory impairment
• Verbal skill to interact	• Significant sensory impairment
• Participant or Legal decision-maker informed consent	

Participants

We recruited 17 participants (avg age: 77.7 years; std = 6.43 years; 15 female) in our study. The participants lived in different senior care facilities such as Kavod, Eaton, Christian Living, Granville, and Cherry Creek senior homes. Table 6.5 contains the including and exclusion criteria.

Study design

Each participant in the study had Ryan placed in his/her apartment for 6 to 10 weeks. They could interact or play games with Ryan at any time. We trained each participant on how to access and use the games. We also created a (hard copy) user manual for the participant. The user manual contained general information on every feature of Ryan including the objective and description of all the games and how to play them.

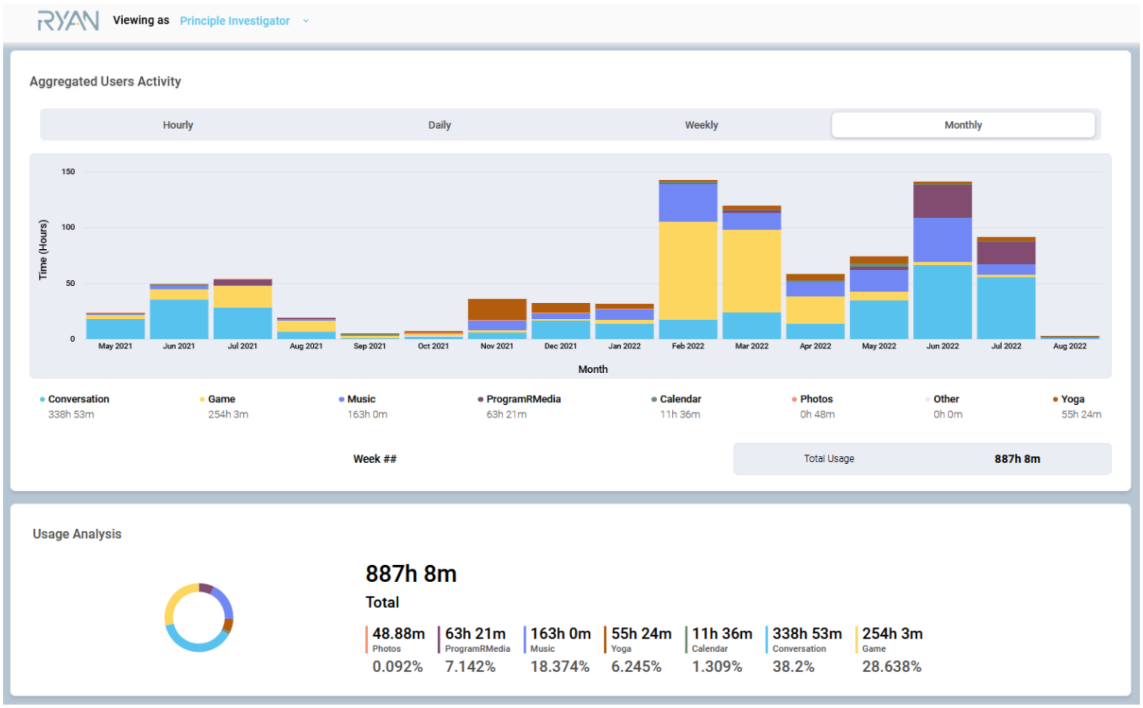


Figure 6.10: The total usage time of Ryan by all the study participants is 887 hours.

Measurement

The type and duration of all activities with Ryan (such as conversation or game) are recorded. We also collect statistics for each game, including usage time, win/loss rate (if applicable), difficulty level, average score, and average duration per game, etc. We also measured PHQ-9, and SLUMS scores, pre, and post-study. The duration of the study was 8-10 weeks. At the end of the study, we conducted an exit survey.

Results

As the Figure 6.10 shows, in total, participants in the study spent 887 hours using Ryan. They spent 254 hours (28.7%) playing with Ryan. This means that on average each user spent 12 minutes (std = 5 minutes) per day playing brain games with Ryan. Checkers, Flow Game, Picture Puzzle, Solitaire, and Word Puzzle, were the most popular games among all. The results of the survey indicated that users preferred these games because they were familiar. They did not like simpler games as “they were not challenging”.

One of the concerns in the field of socially assistive robotics is *novelty effect*, which refers to the phenomenon in which users may lose interest in the robot as they become more familiar with it. This effect is particularly relevant because it could hinder the adoption and long-term use of social robots.

To investigate whether the novelty effect occurred in this study, we evaluated the usage patterns of Ryan by its users over time. Daily usage data from each user were collected and plotted during the study period. The usage patterns of the users did not show a significant decrease in usage, indicating that the novelty effect was not observed. However, it should be noted that the usage patterns appeared to be periodic, with no specific event driving the peaks and valleys of usage. This finding suggests the possibility that users may seek companionship during specific periods. More research is needed to explore the underlying reasons for the observed periodical usage patterns.

The Figures 6.11 and 6.12, the improvement in the users’ scores for PHQ-9 and SLUMS is presented. These two measures are commonly used to assess depression and cognitive impairment, respectively. The results demonstrate the potential of Ryan, to slow the progression of cognitive decline and alleviate the symptoms of Alzheimer’s disease and related dementias (AD/ADRD).The results show an average improvement of 3.3 points for the SLUMS scores and 2.88 points for the PHQ-9 scores, indicating a significant positive impact on the cognitive and emotional states of the users.

Our study also shows that on average, participants performed better over time in playing the games. As shown in Figure 6.13, the average win rate or scores of participants in playing the Flow Game, Word Puzzle, improved over time. That means that the participants were more engaged in playing these games and were able to learn and improve, which is also evident in their improvement in the SLUMS score.

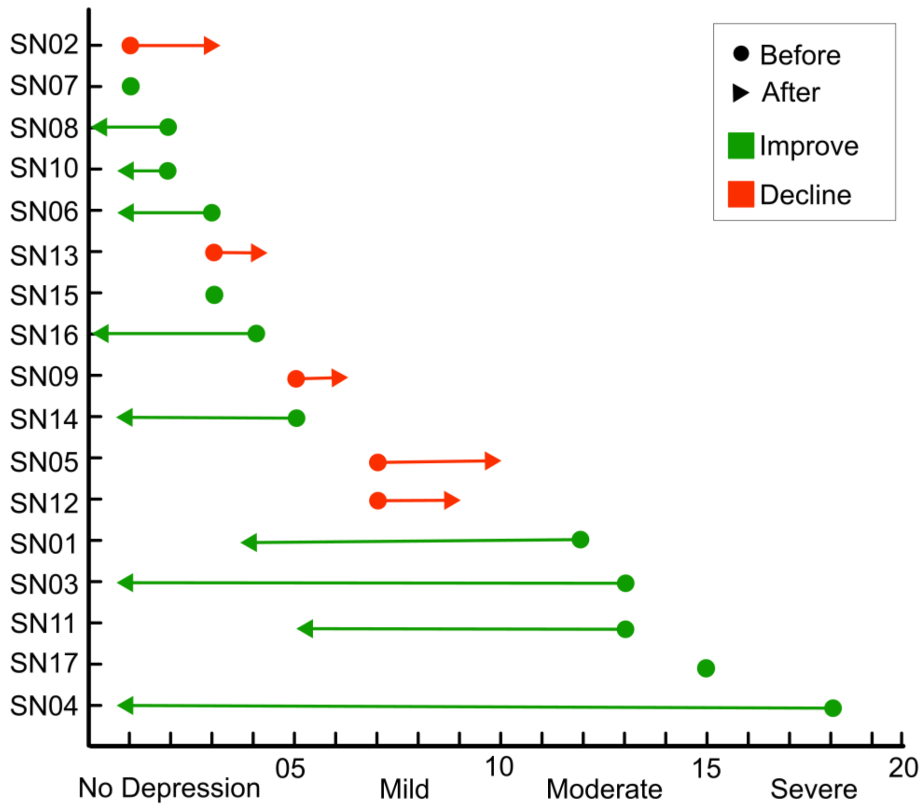


Figure 6.11: PHDQ-9 scores measured before (baseline) and at the end of the study. The subjects are ordered based on their depression score at the start of the study. The subjects with higher depression score benefitted the most.

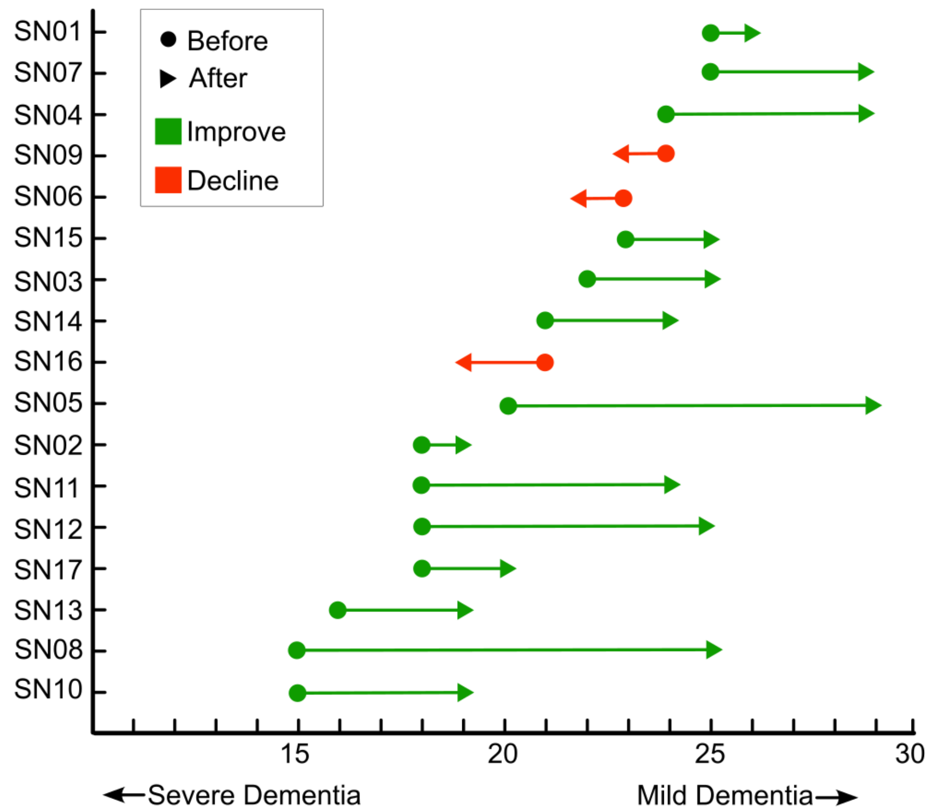


Figure 6.12: SLUMS scores measured before (baseline) and at the end of the study. The subjects are ordered based on their SLUMS score at the start of the study.

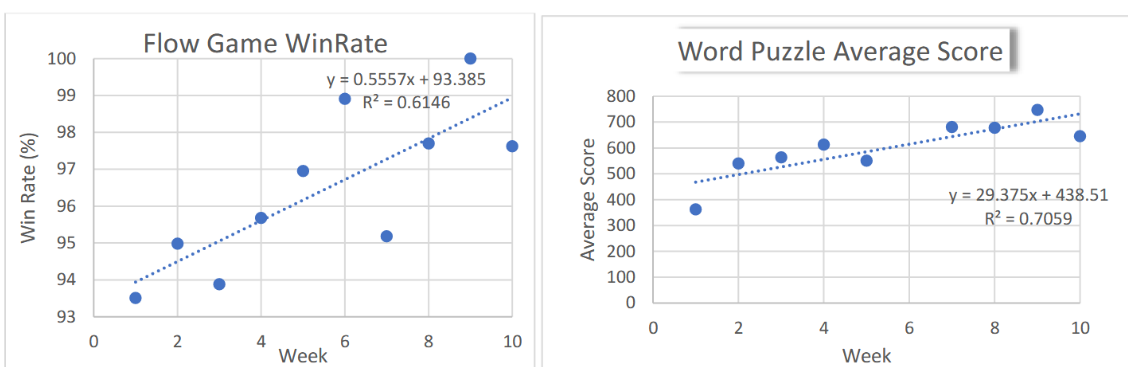


Figure 6.13: Users' improvement in Flow Game and Word Puzzle.

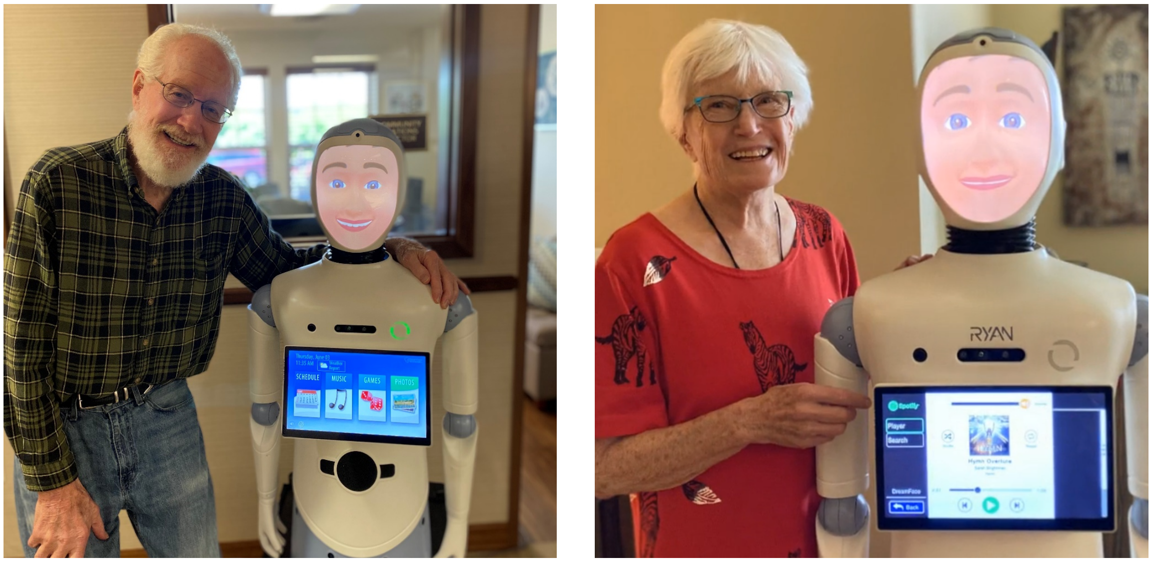


Figure 6.14: The new version of Ryan is emotionally more intelligent.

In the exit survey, the participants gave scores of 4.17 out of 5, to the question "I enjoyed interacting with the robot." Additionally, they gave a score of 4.3 and 4.5 out of 5 to the questions "I find Ryan likable" and "I find Ryan friendly", respectively. See Appendix C for the complete survey.

6.6 Conclusion

The increasing population of the elderly and widespread understaffing in nursing homes can worsen residents' feelings of loneliness and overload nursing staff. During the COVID-19 pandemic, this problem became more evident Xu et al. (2020). The development of AI technologies drew attention to service robots and SAR as potential solutions to these problems. Robots can effectively relieve the burden on healthcare workers and improve the well-being of elderly individuals. Such robots need to be socio-emotionally intelligent in order to effectively engage the aging population.

In this chapter, we discuss a socially assistive robot and its multimodal emotion recognition and multimodal emotion expression systems. More specifically, we compared two versions of the robot: one that uses a scripted dialogue that does not factor in the users' emotions and is lacking facial expressions (non-empathic version), and one with facial expressions that uses an affective dialogue manager to generate a response and has the capability to recognize users' emotions (empathic version).

We studied the differences and effects of Ryan's two versions with a cohort of older adults living in a senior care facility. The statistical analysis of the users' face-scale mood measurement (illustrated in Figure 6.9) indicates an overall positive effect as a result of the interaction with Ryan, irrespective of the robot being empathic or non-empathic. However, word count measurement (Table 6.3) and exit survey analyzes (Table 2) suggest that empathic Ryan is perceived as more engaging and likable. Considering that the only difference between Ryan's two versions is empathic versus non-empathic, the findings suggest that empathy can encourage users to have longer conversations. However, more experiments are needed to further study interactions using a more natural dialogue manager (chatbot). Changes in user depression measurement scores (Table 6.4) suggest that Ryan can potentially decrease users' depression, although to verify this finding, more subjects and long-term studies are required.

We then conducted longitudinal research investigating Ryan's potential to slow the progression of cognitive decline and reduce the symptoms of AD/ADRD. Our approach prioritized human-robot social interaction, companionship, and cognitive game play. The participants played serious brain games with Ryan including Picture Puzzle, Word Puzzle, Chess, Checkers, etc. Our study shows average improvements in SLUMS and PHQ-9 of 3.3 points and 2.88 points, respectively.

Chapter 7

Conclusion, Limitations, and Future works

This dissertation presented the design, development, and successful integration of a social robot equipped with artificial emotional intelligence. This robot is created to improve the quality of life of older adults with dementia and depression. Due to the increasing population of the older adults and the widespread understaffing in nursing homes it is important to study the feasibility of using a social robot as a companion, to improve the quality of life of older adults. During the COVID-19 pandemic, this problem became more evident Xu et al. (2020) as visiting senior care facilities became impossible. Worsening the isolation of this vulnerable group. The development of AI technologies drew attention to service robots and SAR as potential solutions to these problems. Robots can effectively relieve the burden on healthcare workers and improve the well-being of elderly individuals. Such robots need to be socio-emotionally intelligent in order to effectively engage the aging population.

We studied the effects of the artificially emotionally intelligent Ryan on a cohort of older adults living in a senior care facility. The statistical analysis of the results indicates an overall positive effect as a result of the interaction with Ryan and an improvement to both PHQ-9 and SLUMS. While there has been progress in the users' scores, it is crucial to carry out additional studies that involve a larger number of participants and a control group to establish the validity of these findings. Especially, besides interaction with Ryan, other factors such as participants' medical, personal, and environmental conditions should be considered as they can influence their quality of life.

7.1 Ethical Consideration

The use of socially assistive robots in healthcare, particularly elder care, raises ethical considerations that must be carefully evaluated. On the one hand, these robots have the potential to address issues related to loneliness and social isolation, which are prevalent among older adults. By providing companionship and emotional support, socially assistive robots can improve the quality of life of those who may not have regular access to human interaction. However, their use also raises questions about privacy, autonomy, and the potential for overreliance on technology. It is important to ensure that these robots are designed and used in a way that respects individuals' dignity, privacy, and autonomy, and that they are not viewed as a substitute for human companionship and care. In this dissertation, all the studies were conducted with proper Institutional Review Board (IRB) approval. By conducting the research with proper IRB approval, we ensured that the study was conducted in an ethical and responsible manner. Furthermore, the participants were explicitly informed that Ryan is a machine, and they were advised not to take offense in case of any impoliteness. The research conductor provided instructions to the participants to report any such instances to update the dialog manager accordingly.

7.2 Future Work

While this dissertation has demonstrated the potential of the proposed emotionally intelligent robot, there are some future research and improvements that can be made to enhance the proposed platform. Some of these directions are:

1. **Improving the accuracy of the facial expression recognition system:** The proposed FER system in this dissertation showed a good accuracy in classifying facial images in a controlled environment. However, in practice, the robot is used in different environment with different lighting. The subjects are mainly older adults and their facial expression could be improved using a different dataset or training a more robust algorithm. I am going to create a more accurate and specialized FER for older adults.

2. **Affect recognition from other modalities:** In this dissertation, we only used facial expression recognition to understand user's affect. Although facial expressions play a vital role in social interaction and they are a common nonverbal channel through which an AI systems can recognize humans' internal emotions, human affect sensing can be obtained from a broader range of behavioral cues and signals such as body gestures, head movements, speech acoustic analysis, dialog sentiment analysis. Using multi-modal affect recognition with audiovisual affect sensing and tactile sensors (e.g., heart rate, skin conductivity, thermal signals etc.) can enable social robots to understand non-visible user's affect beyond basic expressions with higher accuracy. I will investigate other methods and try to incorporate them in my moodometer.

3. **Creating a better emotional model:** The emotional model used in these studies (moodometer) is simplistic and not natural. I intend to study other emotional models and improve the moodometer to better represent the user's true emotions.

4. **Including more modalities for emotion:** When it comes to Ryan's perception and sensory input, acoustic signals such as tone of voice, volume of voice and other modalities such as eye movement, gaze, head and body gesture, posture, and even breathing rhythm can be used to determine users' emotional state. Currently, Ryan does not utilize these sensory inputs. Adding these features would make the recognition of users' emotional state and intention more accurate and reliable.

5. **Exploring large language models:** The recent proliferation of chatbots and dialog systems, such as ChatGPT, has shown the power of large language models to have longer and deeper conversations. This is a great opportunity to study the perceived emotional intelligence of a robot equipped with such a sophisticated dialog manager.

Bibliography

Alzheimer's association. https://www.alz.org/alzheimers-dementia/facts-figures. Accessed: 2023-01-13. 2

A. Abbott. Dementia: a problem for our age. *Nature*, 475(7355):S2–S4, 2011. 45

H. Abdollahi, A. Mollahosseini, J. T. Lane, and M. Mahoor. A pilot study on using an intelligent life-like robot as a companion for elderly individuals with dementia and depression. In *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, pages 541–546, Nov 2017. doi: 10.1109/HUMANOIDS.2017.8246925. 60, 61

A. Adams, J. Beer, X. Wu, J. Komsky, and J. Zamer. Social Activities in Community Settings: Impact of COVID-19 and Technology Solutions. *Innovation in Aging*, 4(Suppl 1):957, Dec. 2020. ISSN 2399-5300. doi: 10.1093/geroni/igaa057.3499. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7741293/. 58

S. Al Moubayed and G. Skantze. Perception of gaze direction for situated interaction. In *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction*, page 3. ACM, 2012. 27, 28

T. Allison, A. Puce, and G. McCarthy. Social perception from visual cues: role of the sts region. *Trends in cognitive sciences*, 4(7):267–278, 2000. 26

M. Álvarez, R. Galán, F. Matía, D. Rodríguez-Losada, and A. Jiménez. An emotional model for a guide robot. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 40(5):982–992, 2010. 62, 65

P. Alves-Oliveira, P. Sequeira, F. S. Melo, G. Castellano, and A. Paiva. Empathic robot for group learning: A field study. *ACM Transactions on Human-Robot Interaction (THRI)*, 8(1):1–34, 2019. 11, 62

S. M. Anstis, J. W. Mayhew, and T. Morley. The perception of where a face or television'portrait'is looking. *The American journal of psychology*, 82(4):474–489, 1969. 27

K. Aung, M. S. Nurumal, and W. Bukhari. Loneliness among elderly in nursing homes. *International Journal for Studies on Children, Women, Elderly And Disabled*, 2:72–8, 2017. 57

Aylien. Aylien text api. URL http://aylien.com/text-api. 15

E. Bagheri, O. Roesler, H.-L. Cao, and B. Vanderborght. A reinforcement learning based cognitive empathy framework for social robots. *International Journal of Social Robotics*, 13(5):1079–1093, 2021. 58

D. Banerjee, M. Rai, et al. Social isolation in covid-19: The impact of loneliness. *International Journal of Social Psychiatry*, 66(6):525–527, 2020. 57

S. Baron-Cohen, R. Campbell, A. Karmiloff-Smith, J. Grant, and J. Walker. Are children with autism blind to the mentalistic significance of the eyes? *British Journal of Developmental Psychology*, 13(4):379–398, 1995. 26

D. Bertero, F. B. Siddique, C.-S. Wu, Y. Wan, R. H. Y. Chan, and P. Fung. Real-time speech emotion and sentiment recognition for interactive dialogue systems. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1042–1047, 2016. 73

N. Bianchi-Berthouze and A. Kleinsmith. A categorical approach to affective gesture recognition. *Connection science*, 15(4):259–269, 2003. 64

I. P. Bodala, N. Churamani, and H. Gunes. Teleoperated robot coaching for mindfulness training: A longitudinal study. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 939–944. IEEE, 2021. 10

M. A. Brackett, S. E. Rivers, and P. Salovey. Emotional intelligence: Implications for personal, social, academic, and workplace success. *Social and Personality Psychology Compass*, 5, 2011a. ISSN 17519004. 61

M. A. Brackett, S. E. Rivers, and P. Salovey. Emotional intelligence: Implications for personal, social, academic, and workplace success. *Social and Personality Psychology Compass*, 5(1):88–103, 2011b. 63

C. Breazeal. Socially intelligent robots. *interactions*, 12(2):19–22, 2005. 4

C. Breazeal, K. Dautenhahn, and T. Kanda. Social robotics. In *Springer handbook of robotics*, pages 1935–1972. Springer, 2016. 4

C. L. Breazeal. *Sociable machines: Expressive social exchange between humans and robots*. PhD thesis, Massachusetts Institute of Technology, 2000. 10

F. Burkhardt, M. V. Ballegooy, K.-P. Engelbrecht, T. Polzehl, and J. Stegmann. Emotion detection in dialog systems: Applications, strategies and challenges. *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–6, 2009. 73

J. Cassell. *Embodied conversational agents*. MIT press, 2000. 26

G. Castellano, R. Aylett, K. Dautenhahn, A. Paiva, P. W. McOwan, and S. Ho. Long-term affect sensitive and socially interactive companions. In *Proceedings of the 4th International Workshop on Human-Computer Conversation*, 2008a. 13

G. Castellano, L. Kessous, and G. Caridakis. Emotion recognition through multiple modalities: face, body gesture, speech. In *Affect and emotion in human-computer interaction*, pages 92–103. Springer, 2008b. 63

F. Cavallo, M. Aquilano, M. Bonaccorsi, R. Limosani, A. Manzi, M. C. Carrozza, and P. Dario. Improving domiciliary robotic services by integrating the astro robot in an ami infrastructure. In *Gearing up and accelerating cross-fertilization between academic and industrial robotics research in Europe: Technology transfer experiments from the ECHORD project*, pages 267–282. Springer, 2014. 8

F. Cavallo, F. Semeraro, L. Fiorini, G. Magyar, P. Sinčák, and P. Dario. Emotion modelling for social robotics applications: a review. *Journal of Bionic Engineering*, 15(2):185–203, 2018. 67

N. Céspedes, B. Irfan, E. Senft, C. A. Cifuentes, L. F. Gutierrez, M. Rincon-Roncancio, T. Belpaeme, and M. Múnera. A socially assistive robot for long-term cardiac rehabilitation in the real world. *Frontiers in Neurorobotics*, 15:633248, 2021a. 10

N. Céspedes, D. Raigoso, M. Múnera, and C. A. Cifuentes. Long-term social human-robot interaction for neurorehabilitation: robots as a tool to support gait therapy in the pandemic. *Frontiers in Neurorobotics*, 15:612034, 2021b. 10

S.-C. Chen, W. Moyle, C. Jones, and H. Petsky. A social robot intervention on depression, loneliness, and quality of life for taiwanese older adults in long-term care. *International psychogeriatrics*, 32(8):981–991, 2020. 58

Y.-C. Chen and S.-L. Yeh. Look into my eyes and i will see you: Unconscious processing of human gaze. *Consciousness and cognition*, 21(4):1703–1710, 2012. 26

M. G. Cline. The perception of where a person is looking. *The American journal of psychology*, 80(1):41–50, 1967. 29

D. Conti, G. Trubia, S. Buono, S. Di Nuovo, and A. Di Nuovo. Evaluation of a robot-assisted therapy for children with autism and intellectual disability. In *Towards Autonomous Robotic Systems: 19th Annual Conference, TAROS 2018, Bristol, UK July 25-27, 2018, Proceedings 19*, pages 405–415. Springer, 2018. 8

R. Cowie and E. Douglas-Cowie. Automatic statistical analysis of the signal and prosodic signs of emotion in speech. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, volume 3, pages 1989–1992. IEEE, 1996. 64

L. J. Cronbach. Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3): 297–334, 1951. 52

J. Cummings. Lessons learned from alzheimer disease: clinical trials with negative outcomes. *Clinical and translational science*, 11(2):147, 2018. 2

A. Damasio. Descartes' error: Emotion, reason and the human brain (new york: Gossett/putnam). 344 christine tappolet d'arms, j. and jacobson, d.(2003), the significance of recalcitrant emotions; or anti-quasi judgmentalism. philosophy. suppl. vol. *Proceedings of the Royal Institute of Philosophy, 127*, 46:287–310, 1994. 4

M. Dapretto, M. S. Davies, J. H. Pfeifer, A. A. Scott, M. Sigman, S. Y. Bookheimer, and M. Iacoboni. Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature neuroscience*, 9(1):28, 2006. 65

B. Dasgupta and T. Mruthyunjaya. The stewart platform manipulator: a review. *Mechanism and machine theory*, 35(1):15–40, 2000. 19

K. Dautenhahn. Socially intelligent robots: dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):679–704, 2007. 4

M. H. Davis. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, 44(1):113, 1983. 77

F. Delaunay, J. de Greeff, and T. Belpaeme. A study of a retro-projected robotic face and its effectiveness for gaze reading by humans. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pages 39–44. IEEE Press, 2010. 27, 39

E. Deng, B. Mutlu, M. J. Mataric, et al. Embodiment in socially interactive robots. *Foundations and Trends® in Robotics*, 7(4):251–356, 2019. 75

D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet, et al. Simsensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1061–1068, 2014. 75

F. Dino, R. Zandie, H. Abdollahi, S. Schoeder, and M. Mahoor. Delivering cognitive behavioral therapy using a conversational social robot. In *Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on*, 2019. 63, 66

DreamFace-Tech. Social robotics, 2015. URL http://dreamfacetech.com/. last checked: 01.20.2017. 65, 66, 75

B. R. Duffy. Anthropomorphism and the social robot. *Robotics and autonomous systems*, 42(3-4):177–190, 2003. 10

B. R. Duffy, G. Joue, and J. Bourke. Issues in assessing performance of social robots. In *WSEAS International Conference on Robotics*, 2002. 10

G. D'Onofrio, D. Sancarlo, J. Oscar, F. Ricciardi, D. Casey, K. Murphy, F. Giuliani, and A. Greco. A multicenter survey about companion robot acceptability in caregivers of patients with dementia. In *Sensors and Microsystems: Proceedings of the 19th AISEM 2017 National Conference 19*, pages 161–178. Springer, 2018. 8

Eaton. Eaton senior communities. http://www.eatonsenior.org/. Accessed: 2023-03-09. 47

P. Ekman and W. Friesen. *The Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, 1978. 67

N. J. Emery. The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, 24(6):581–604, 2000. 26

D. Feil-Seifer and M. J. Mataric. Defining socially assistive robotics. In *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on*, pages 465–468. IEEE, 2005. 44, 57

R. Feingold Polak and S. L. Tzedek. Social robot for rehabilitation: expert clinicians and post-stroke patients' evaluation following a long-term intervention. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*, pages 151–160, 2020. 11

Y. Fernaeus, M. Håkansson, M. Jacobsson, and S. Ljungblad. *How do you play with a robotic toy animal? A long-term study of Pleo*, page 39–48. ACM Press, 2010. ISBN 9781605589510. 13

J. M. Garrido, E. Bofias, Y. Laplaza, M. Marquina, M. Aylett, and C. Pidcock. The cerevoice speech synthesiser. *Actas de las V Jornadas de Tecnología del Habla*, pages 126–129, 2008. 20

M. Ghafurian, C. Ellard, and K. Dautenhahn. Social companion robots to reduce isolation: a perception change due to covid-19. In *IFIP Conference on Human-Computer Interaction*, pages 43–63. Springer, 2021. 57

I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, et al. Challenges in representation learning: A report on three machine learning contests. *Neural Networks*, 64:59–63, 2015. 40

P. E. Greenberg, A.-A. Fournier, T. Sisitsky, M. Simes, R. Berman, S. H. Koenigsberg, and R. C. Kessler. The economic burden of adults with major depressive disorder in the united states (2010 and 2018). *Pharmacoeconomics*, 39(6):653–665, 2021. 3

B. Hasani and M. H. Mahoor. Facial affect estimation in the wild using deep residual and convolutional networks. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, pages 1955–1962. IEEE, 2017. 41

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 41

L. E. Hebert, J. Weuve, P. A. Scherr, and D. A. Evans. Alzheimer disease in the united states (2010–2050) estimated using the 2010 census. *Neurology*, 80(19):1778–1783, 2013. 2

M. Heerink, B. Kröse, B. Wielinga, V. Evers, et al. Studying the acceptance of a robotic agent by elderly users. *International Journal of Assistive Robotics and Mechatronics*, 7 (3):33–43, 2006. 9, 46

A. P. Henkel, M. Čaić, M. Blaurock, and M. Okan. Robotic transformative service research: deploying social robots for consumer well-being during covid-19 and beyond. *Journal of Service Management*, 2020. 58

A. Henschel and E. Cross. The neuroscience of loneliness–and how technology is helping us [internet]. conversat. 2020, 2020. 11

A. Henschel, G. Laban, and E. S. Cross. What makes a robot social? a review of social robots from science fiction to a home or hospital near you. *Current Robotics Reports*, 2: 9–19, 2021. 11

S. Hershey, S. Chaudhuri, D. P. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, et al. Cnn architectures for large-scale audio classification. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pages 131–135. IEEE, 2017. 41

U. Hess and S. Blairy. Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International journal of psychophysiology*, 40(2):129–141, 2001. 65

D. Heylen. Head gestures, gaze and the principles of conversational structure. *International Journal of Humanoid Robotics*, 3(03):241–267, 2006. 18

J. Hill, W. R. Ford, and I. G. Farreras. Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in human behavior*, 49:245–250, 2015. 77

M. Imai, T. Kanda, T. Ono, H. Ishiguro, and K. Mase. Robot mediated round table: Analysis of the effect of robot's gaze. In *Robot and Human Interactive Communication, 2002. Proceedings. 11th IEEE International Workshop on*, pages 411–416. IEEE, 2002. 26

B. Irfan, A. Narayanan, and J. Kennedy. Dynamic emotional language adaptation in multiparty interactions with agents. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, pages 1–8, 2020. 75

R. J. Itier and M. Batty. Neural bases of eye and gaze processing: the core of social cognition. *Neuroscience & Biobehavioral Reviews*, 33(6):843–863, 2009. 26

K. S. Judge, C. J. Camp, and S. Orsulic-Jeras. Use of montessori-based activities for clients with dementia in adult day care: Effects on engagement. *American Journal of Alzheimer's Disease*, 15(1):42–46, 2000. 16

M. Kanamori, M. Suzuki, H. Oshiro, M. Tanaka, T. Inoguchi, H. Takasugi, Y. Saito, and T. Yokoyama. Pilot study on improvement of quality of life among elderly using a pet-type robot. In *Proceedings 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation. Computational Intelligence in Robotics and Automation for the New Millennium (Cat. No. 03EX694)*, volume 1, pages 107–112. IEEE, 2003. 66

A. H. Kargar B and M. H. Mahoor. A Pilot Study on the eBear Socially Assistive Robot: Implication for Interacting with Elderly People with Moderate Depression. In *IEEE-RAS International Conference on Humanoid Robots*, Birmingham, UK, Nov. 2017. 77, 80

Z. Kasap, M. B. Moussa, P. Chaudhuri, and N. Magnenat-Thalmann. Making them remember—emotional virtual characters with memory. *IEEE Computer Graphics and Applications*, 29(2):20–29, 2009. 75

A. Kendon. Some functions of gaze-direction in social interaction. *Acta psychologica*, 26: 22–63, 1967. 26

R. Khosla, M.-T. Chu, R. Kachouie, K. Yamada, F. Yoshihiro, and T. 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, 2012. 8

C. D. Kidd, W. Taggart, and S. Turkle. A sociable robot to encourage social interaction among the elderly. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 3972–3976. IEEE, 2006. 12

Kinect. Kinect for windows sdk. http://kinectforwindows.org. Accessed: 2017-10-05. 15

M. Kinya and E. Mitsuo. Illusory face dislocation effect and configurational integration in the inverted face. *Tohoku Psychologica Folia*, 43(1-4):150–160, 1984. 29

N. L. Kluttz, B. R. Mayes, R. W. West, and D. S. Kerby. The effect of head turn on the perception of gaze. *Vision research*, 49(15):1979–1993, 2009. 29

A. A. Kotwal, J. Kim, L. Waite, and W. Dale. Social function and cognitive status: Results from a us nationally representative survey of older adults. *Journal of General Internal Medicine*, 31(8):854–862, Apr 2016. ISSN 1525-1497. doi: 10.1007/s11606-016-3696-0. URL http://dx.doi.org/10.1007/s11606-016-3696-0. 8, 46, 66

N. C. Krämer, A. von der Pütten, and S. Eimler. Human-agent and human-robot interaction theory: similarities to and differences from human-human interaction. In *Human-computer interaction: The agency perspective*, pages 215–240. Springer, 2012. 9, 46

K. Kroencke, R. Spitzer, and J. Williams. The phq-9: validity of a brief depression severity measure [electronic version]. *Journal of General Internal Medicine*, 16(9):606–13, 2001. 48, 82, 83

S. R. Langton, H. Honeyman, and E. Tessler. The influence of head contour and nose angle on the perception of eye-gaze direction. *Perception & psychophysics*, 66(5):752–771, 2004. 29

V. Lawrence, J. Fossey, C. Ballard, E. Moniz-Cook, and J. Murray. Improving quality of life for people with dementia in care homes: making psychosocial interventions work. *The British Journal of Psychiatry*, 201(5):344–351, 2012. 16, 17

K. M. Lee, N. Park, and H. Song. Can a robot be perceived as a developing creature? *Human Communication Research*, 31(4):538–563, Oct 2005. ISSN 0360-3989. 9

I. Leite. Long-term interactions with empathic social robots. *AI Matters*, 1(3):13–15, 2015. 8, 46

I. Leite, S. Mascarenhas, C. Martinho, R. Prada, and A. Paiva. “why can’t we be friends?” an empathic game companion for long-term interaction. page 315–321, 2010. 44

I. Leite, C. Martinho, and A. Paiva. Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, 5:291–308, 2013a. 10

I. Leite, A. Pereira, S. Mascarenhas, C. Martinho, R. Prada, and A. Paiva. The influence of empathy in human–robot relations. *International journal of human-computer studies*, 71(3):250–260, 2013b. 11, 62

C. Linnemann and U. E. Lang. Pathways connecting late-life depression and dementia. *Frontiers in Pharmacology*, 11:279, 2020. ISSN 1663-9812. doi: 10.3389/fphar.2020.00279. URL https://www.frontiersin.org/article/10.3389/fphar.2020.00279. 60

J. X. Liu, Y. Goryakin, A. Maeda, T. Bruckner, and R. Scheffler. Global health workforce labor market projections for 2030. *Human resources for health*, 15(1):1–12, 2017. 3

S. M. Loi, A. Bennett, M. Pearce, K. Nguyen, N. T. Lautenschlager, R. Khosla, and D. Velakoulis. A pilot study exploring staff acceptability of a socially assistive robot in a residential care facility that accommodates people under 65 years old. *International psychogeriatrics*, 30(7):1075–1080, 2018. 8

C. D. Lorish and R. Maisiak. The face scale: a brief, nonverbal method for assessing patient mood. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, 29(7):906–909, 1986. 77, 80

C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60, 2014. URL http://www.aclweb.org/anthology/P/P14/P14-5010. 70

P. Marti, M. Bacigalupo, L. Giusti, C. Mennecozzi, and T. Shibata. Socially assistive robotics in the treatment of behavioural and psychological symptoms of dementia. In *Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on*, pages 483–488. IEEE, 2006. 45

D. McDuff and M. Czerwinski. Designing emotionally sentient agents. *Communications of the ACM*, 61(12):74–83, 2018. 62, 70

J. E. Michaelis and B. Mutlu. Someone to read with: Design of and experiences with an in-home learning companion robot for reading. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 301–312, 2017. 10

M. Minsky. *Society of mind*. Simon and Schuster, 1988. 4

K. Misawa, Y. Ishiguro, and J. Rekimoto. Livemask: A telepresence surrogate system with a face-shaped screen for supporting nonverbal communication. In *Proceedings of the international working conference on advanced visual interfaces*, pages 394–397. ACM, 2012. 27, 28, 38

A. Mollahosseini, G. Graitzer, E. Borts, S. Conyers, R. M. Voyles, R. Cole, and M. H. Mahoor. *ExpressionBot: An emotive lifelike robotic face for face-to-face communication*, page 1098–1103. IEEE, Nov 2014a. ISBN 978-1-4799-7174-9. 12

A. Mollahosseini, G. Graitzer, E. Borts, S. Conyers, R. M. Voyles, R. Cole, and M. H. Mahoor. Expressionbot: An emotive lifelike robotic face for face-to-face communication. In *Humanoid Robots (Humanoids), 2014 14th IEEE-RAS International Conference on*, pages 1098–1103. IEEE, 2014b. 14, 27, 28, 38

A. Mollahosseini, B. Hasani, M. J. Salvador, H. Abdollahi, D. Chan, and M. H. Mahoor. Facial expression recognition from world wild web. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016. 40

A. Mollahosseini, B. Hasani, and M. H. Mahoor. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, PP(99):1–1, 2017. ISSN 1949-3045. doi: 10.1109/TAFFC.2017.2740923. 20, 40, 68

A. Mollahosseini, H. Abdollahi, and M. H. Mahoor. Studying effects of incorporating automated affect perception with spoken dialog in social robots. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 783–789. IEEE, 2018a. 11, 39, 43, 62

A. Mollahosseini, H. Abdollahi, T. Sweeny, R. Cole, and M. Mahoor. Role of embodiment and presence in human perception of robots' facial cues. *International Journal of Human-Computer Studies*, 116:25–39, 2018b. 25, 75

E. Mordoch, A. Osterreicher, L. Guse, K. Roger, and G. Thompson. Use of social commitment robots in the care of elderly people with dementia: A literature review. *Maturitas*, 74(1):14–20, 2013. 11, 47

S. A. Moubayed, J. Edlund, and J. Beskow. Taming mona lisa: Communicating gaze faithfully in 2d and 3d facial projections. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 1(2):11, 2012. 27, 28, 39

B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, and N. Hagita. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 61–68. ACM, 2009. 26

NAO. Nao. https://www.softbankrobotics.com/emea/en/robots/nao. Accessed: 2018-09-13. 63, 65

R. Nielsen, R. Voyles, D. Bolaños, M. Mahoor, W. Pace, K. Siek, and W. Ward. A platform for human-robot dialog systems research. In *AAAI Fall Symposium: Dialog with Robots*, pages 161–162, 2010. 59

NIMH. Major depression, nimh. https://www.nimh.nih.gov/health/statistics/major-depression, 2020. Accessed: 2023-02-13. 3

T. Nomura, T. Kanda, T. Suzuki, and S. Yamada. Do people with social anxiety feel anxious about interacting with a robot? *Ai & Society*, 35:381–390, 2020. 11

T. L. Nwe, S. W. Foo, and L. C. D. Silva. Speech emotion recognition using hidden markov models. *Speech Communication*, 41:603–623, 2003. 73

M. Ochs, R. Niewiadomski, C. Pelachaud, and D. Sadek. Intelligent expressions of emotions. In *International Conference on Affective Computing and Intelligent Interaction*, pages 707–714. Springer, 2005. 61

Y. Otsuka, I. Mareschal, A. J. Calder, and C. W. Clifford. Dual-route model of the effect of head orientation on perceived gaze direction. *Journal of Experimental Psychology: Human perception and performance*, 40(4):1425, 2014. 29

A. Paiva, J. Dias, D. Sobral, R. Aylett, S. Woods, L. Hall, and C. Zoll. Learning by feeling: Evoking empathy with synthetic characters. *Applied Artificial Intelligence*, 19(3-4):235–266, 2005. 11, 62

Pandorabots. Pandorabots. URL http://www.pandorabots.com/. 16

M. Pantic, N. Sebe, J. F. Cohn, and T. Huang. Affective multimodal human-computer interaction. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 669–676. ACM, 2005. 63

N. Pateromichelakis, A. Mazel, M. Hache, T. Koumpogiannis, R. Gelin, B. Maisonnier, and A. Berthoz. Head-eyes system and gaze analysis of the humanoid robot romeo. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 1374–1379. IEEE, 2014. 26

Pepper. Pepper robot. https://www.softbankrobotics.com/emea/en/pepper/. [Online; accessed 19-July-2021]. 5, 8, 65

D. Perrett, P. Smith, D. Potter, A. Mistlin, A. Head, A. Milner, and M. Jeeves. Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London B: Biological Sciences*, 223(1232):293–317, 1985. 26

S. Petersen, S. Houston, H. Qin, C. Tague, and J. Studley. The utilization of robotic pets in dementia care. *Journal of Alzheimer's Disease*, 55(2):569–574, 2017. 63

R. W. Picard. *Affective computing*. MIT press, 2000. 59, 61, 62

J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun. Towards robotic assistants in nursing homes: Challenges and results. *Robotics and Autonomous Systems*, 42(3-4): 271–281, Mar 2003. ISSN 0921-8890. doi: 10.1016/s0921-8890(02)00381-0. URL http://dx.doi.org/10.1016/S0921-8890(02)00381-0. 10, 47

J. W. Plunkett. *Plunkett's Health Care Industry Almanac 2007: Health Care Industry Market Research, Statistics, Trends & Leading Companies*. Plunkett Research, Ltd. 2

D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, et al. The kaldi speech recognition toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding*, number CONF. IEEE Signal Processing Society, 2011. 20

H. Prendinger and M. Ishizuka. The empathic companion: A character-based interface that addresses users'affective states. *Applied Artificial Intelligence*, 19(3-4):267–285, 2005. 58, 64

S. D. Preston and F. B. De Waal. Empathy: Its ultimate and proximate bases. *Behavioral and brain sciences*, 25(1):1–20, 2002. 42, 62

Program-Y. Program-y. https://github.com/keiffster/program-y. 72

T. E. Project. The emote project, 2013. URL http://gaips.inesc-id.pt/emote/. last checked: 01.20.2017. 77

L. Pu, W. Moyle, C. Jones, and M. Todorovic. The effectiveness of social robots for older adults: a systematic review and meta-analysis of randomized controlled studies. *The Gerontologist*, 59(1):e37–e51, 2019. 58

M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, A. Y. Ng, et al. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, page 5. Kobe, Japan, 2009. 20

S. M. Rabbitt, A. E. Kazdin, and B. Scassellati. Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. *Clinical Psychology Review*, 35:35–46, 2015. 44

S. Rasouli, G. Gupta, E. Nilsen, and K. Dautenhahn. Potential applications of social robots in robot-assisted interventions for social anxiety. *International Journal of Social Robotics*, 14(5):1–32, 2022. 11

Realsense. Intel realsense sdk for windows. https://software.intel.com/en-us/intel-realsense-sdk. Accessed: 2017-10-05. 15

S. Ren, X. Cao, Y. Wei, and J. Sun. Face alignment at 3000 fps via regressing local binary features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1685–1692, 2014. 41

RestfulAPI. Restful api. https://restfulapi.net/. Accessed: 2018-09-13. 74

N. L. Robinson, T. V. Cottier, and D. J. Kavanagh. Psychosocial health interventions by social robots: systematic review of randomized controlled trials. *Journal of medical Internet research*, 21(5):e13203, 2019. 11

N. L. Robinson, J. Connolly, L. Hides, and D. J. Kavanagh. A social robot to deliver an 8-week intervention for diabetes management: Initial test of feasibility in a hospital clinic. In *Social Robotics: 12th International Conference, ICSR 2020, Golden, CO, USA, November 14–18, 2020, Proceedings 12*, pages 628–639. Springer, 2020. 10

D. M. Romano, G. Sheppard, J. Hall, A. Miller, and Z. Ma. Basic: A believable, adaptable socially intelligent character for social presence. In *Proceedings of the 8th Annual International Workshop on Presence*, pages 287–290. Citeseer, 2005. 75

K. Ruhland, S. Andrist, J. Badler, C. Peters, N. Badler, M. Gleicher, B. Mutlu, and R. Mcdonnell. Look me in the eyes: A survey of eye and gaze animation for virtual agents and artificial systems. In *Eurographics State-of-the-Art Report*, pages 69–91, 2014. 26

J. Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39 (6):1161–1178, 1980. 67

Ryan. Dreamface-tech., social robotics. URL http://dreamfacetech.com/. last checked: 01.20.2017. 5, 12

D. Sander. Models of emotion: The affective neuroscience approach. *The Cambridge Handbook of Human Affective Neuroscience*, 2013. 67

M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018. 20

M. Sarabia, N. Young, K. Canavan, T. Edginton, Y. Demiris, and M. P. Vizcaychipi. Assistive robotic technology to combat social isolation in acute hospital settings. *International Journal of Social Robotics*, 10(5):607–620, 2018. 63

R. V. Saveanu and C. B. Nemeroff. Etiology of depression: genetic and environmental factors. *Psychiatric clinics*, 35(1):51–71, 2012. 3

B. Scassellati and M. Vázquez. The potential of socially assistive robots during infectious disease outbreaks. *Science Robotics*, 5(44):eabc9014, 2020. 11

M. Schroder, E. Bevacqua, R. Cowie, F. Eyben, H. Gunes, D. Heylen, M. Ter Maat, G. McKeown, S. Pammi, M. Pantic, et al. Building autonomous sensitive artificial listeners. *IEEE transactions on affective computing*, 3(2):165–183, 2011. 75

N. Sebe, I. Cohen, and T. S. Huang. Multimodal emotion recognition. In *Handbook of Pattern Recognition and Computer Vision*, pages 387–409. World Scientific, 2005. 63

W. Shi and Z. Yu. Sentiment adaptive end-to-end dialog systems. *arXiv preprint arXiv:1804.10731*, 2018. 64, 73

Simon. Simon the robot. URL http://www.simontherobot.com/. 12

Socibot. Socibot platform of engineeredarts, 2015. URL https://www.engineeredarts.co.uk/socibot/. last checked: 01.20.2017. 65

Sophia. Sophia, hanson robotics. http://www.hansonrobotics.com/robot/sophia. Accessed: 2020-01-13. 5

M. Spezialetti, G. Placidi, and S. Rossi. Emotion recognition for human-robot interaction: recent advances and future perspectives. *Frontiers in Robotics and AI*, 7, 2020. 63

T. D. Sweeny and D. Whitney. The center of attention: Metamers, sensitivity, and bias in the emergent perception of gaze. *Vision Research*, 131:67–74, 2017. 29

T. D. Sweeny, S. Haroz, and D. Whitney. Reference repulsion in the categorical perception of biological motion. *Vision research*, 64:26–34, 2012. 26

M. Szabóová, M. Sarnovskỳ, V. Maslej Krešňáková, and K. Machová. Emotion analysis in human–robot interaction. *Electronics*, 9(11):1761, 2020. 67

C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015. 68

W. Taggart, S. Turkle, and C. D. Kidd. An interactive robot in a nursing home : Preliminary remarks field setting : Nursing homes. *Towards Social Mechanisms of Android Science: A COGSCI Workshop*, page 1–6, 2005. 44

A. Tapus, M. Maja, and B. Scassellatti. The grand challenges in socially assistive robotics. *IEEE Robotics and Automation Magazine*, 14(1):35–42, 2007. 58

S. Tariq, N. Tumosa, J. Chibnall, H. Perry III, and J. Morley. The saint louis university mental status (slums) examination for detecting mild cognitive impairment and dementia is more sensitive than the mini-mental status examination (mmse)–a pilot study. *Am J Geriatr Psychiatry*, 14(11):900–10, 2006. 48, 76

Y. Tian, T. Kanade, and J. F. Cohn. Facial expression recognition. In *Handbook of face recognition*, pages 487–519. Springer, 2011. 40

D. Todorović. Geometrical basis of perception of gaze direction. *Vision research*, 46(21): 3549–3562, 2006. 26

C.-N. Tseng, B.-S. Gau, and M.-F. Lou. The effectiveness of exercise on improving cognitive function in older people: a systematic review. *Journal of Nursing Research*, 19(2): 119–131, 2011. 17

A. Van Maris, N. Zook, P. Caleb-Solly, M. Studley, A. Winfield, and S. Dogramadzi. Designing ethical social robots—a longitudinal field study with older adults. *Frontiers in Robotics and AI*, 7:1, 2020. 10

T. Vandemeulebroucke, B. D. de Casterlé, and C. Gastmans. How do older adults experience and perceive socially assistive robots in aged care: a systematic review of qualitative evidence. *Aging & mental health*, 22(2):149–167, 2018. 57

L. P. Vardoulakis, L. Ring, B. Barry, C. L. Sidner, and T. Bickmore. Designing relational agents as long term social companions for older adults. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7502 LNAI:289–302, 2012. ISSN 03029743. 10, 46, 47

P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004. 67

K. Wada, T. Shibata, T. Saito, and K. Tanie. Effects of robot assisted activity to elderly people who stay at a health service facility for the aged. In *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 3, pages 2847–2852. IEEE, 2003. 9, 46, 63, 66

R. Wallace. The elements of aiml style. *Alice AI Foundation*, 139, 2003. 73

R. S. Wallace. The anatomy of alice. In *Parsing the Turing Test*, pages 181–210. Springer, 2009. 73

W. H. Wollaston. On the apparent direction of eyes in a portrait. *Philosophical Transactions of the Royal Society of London*, 114:247–256, 1824. 29

R. F. Woolson. Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials*, pages 1–3, 2007. 80

H. Xu, O. Intrator, and J. R. Bowblis. Shortages of staff in nursing homes during the covid-19 pandemic: What are the driving factors? *Journal of the American Medical Directors Association*, 21(10):1371–1377, 2020. 58, 93, 95

G.-Z. Yang, B. J. Nelson, R. R. Murphy, H. Choset, H. Christensen, S. H. Collins, P. Dario, K. Goldberg, K. Ikuta, N. Jacobstein, et al. Combating covid-19—the role of robotics in managing public health and infectious diseases, 2020. 11

J. A. Yesavage and J. I. Sheikh. 9/geriatric depression scale (gds) recent evidence and development of a shorter version. *Clinical gerontologist*, 5(1-2):165–173, 1986. 82, 83

R. Yonck. *Heart of the machine: Our future in a world of artificial emotional intelligence*. Arcade, 2020. 4, 58

Y. Yoshikawa, K. Shinozawa, H. Ishiguro, N. Hagita, and T. Miyamoto. Responsive robot gaze to interaction partner. In *Robotics: Science and systems*, 2006. 26

Z.-J. You, C.-Y. Shen, C.-W. Chang, B.-J. Liu, and G.-D. Chen. A robot as a teaching assistant in an english class. In *Advanced Learning Technologies, 2006. Sixth International Conference on*, pages 87–91. IEEE, 2006. 13

L. Yu. Face alignment in 3000fps. https://github.com/yulequan/face-alignment-in-3000fps, 2016. 41

Z. Yu, A. Papangelis, and A. Rudnicky. Ticktock: A non-goal-oriented multimodal dialog system with engagement awareness. In *2015 AAAI Spring symposium series*, pages 108–111, 2015. 58

M. Zamora-Macorra, E. F. A. de Castro, J. A. Ávila-Funes, B. S. Manrique-Espinoza, R. López-Ridaura, A. L. Sosa-Ortiz, P. L. Shields, and D. S. M. del Campo. The association between social support and cognitive function in mexican adults aged 50 and older. *Archives of Gerontology and Geriatrics*, 68:113–118, 2017. 8, 46, 66

Zeno. Zeno, 2009. URL http://www.hansonrobotics.com/robot/zeno/. 65

Appendices

A Full questionnaire used in the pilot study

Table 1: The mean rank and questions of the exit survey evaluating users' likability and acceptance of interacting with Ryan and its features (1-strongly disagree, 5-strongly agree).

	Question	Avg. Score ± (STD)	Cronbach's alpha
Questions About User Interaction with Ryan	Q1. I enjoyed interacting with the robot.	4.17 ± 0.75	0.930
	Q2. The conversation with the robot was interesting.	4.00 ± 0.89	
	Q3. Talking with the robot was like talking to a person.	3.00 ± 1.54	
	Q4. I feel happier when I had the robot as my company.	3.67 ± 1.03	
	Q5. I would like to have this robot in my home again.	3.33 ± 1.50	
	Q6. I feel less depressed after talking to the robot.	3.33 ± 1.36	

Continued on the next page

Table 1 – Continued from the previous page			
	Question	**Avg. Score ± (STD)**	**Cronbach's alpha**
Questions About Feature of Ryan	Q7. I liked the robot's facial expressions.	4.17 ± 0.75	0.924
	Q8. I liked the robot mirroring my facial expressions.	3.50 ± 1.04	
	Q9. The robot reminder helped me to be on schedule.	4.00 ± 0.63	
	Q10. I enjoyed the robot playing my favorite music.	4.17 ± 0.40	
	Q11. I enjoyed the robot playing videos for me.	3.83 ± 0.75	
	Q12. The videos were effective and affected my life style.	3.50 ± 1.51	
	Q13. I enjoyed playing the games.	3.33 ± 1.50	
	Q14. The games helped me train my brain, though they were simple.	3.17 ± 1.32	
	Q15. The games were challenging.	2.00 ± 1.54	
	Q16. I enjoyed watching my photo album shown by the robot.	4.33 ± 0.81	

B Full survey questionnaire used in AEI study

Table 2: the mean rank and questions of the exit survey evaluating users' likability and Ryan's emotion and sympathy with participants.

	Question	mean±std
Evaluation of Ryan Empathy and Emotion	Q1. I felt Ryan was gentle with me.	4.80±.40
	Q3. On scale of 1-5 how would you rate Ryan's facial expressions?	4.38±.70
	Q4. I feel happier when I was in the company of Ryan.	4.50±.67
	Q5. I feel less depressed after talking to Ryan.	4.50±.81
	Q6. I felt Ryan understood my emotions.	4.10±1.04
	Q7. Ryan encouraged me to open up about my mood/feelings.	4.00±1.18
	Q8. The sessions with Ryan improved my mood and made me feel happier than I was before we began the session.	4.50±.92
	Q10. How much do you agree with the statement: "I like that Ryan empathized with me"	4.57±.49
	Q11. How well did Ryan empathize with your feelings?	4.40±.80
	Q23. I feel happier when I was in the company of Ryan.	4.60± .66

Continued on the next page

Table 2 – Continued from the previous page		
	Question	**mean±std**
Evaluation of the interaction with Ryan and Likeability (Entire study)	Q14. The conversation with Ryan was enjoyable.	4.50±.67
	Q15. The conversation with Ryan was engaging.	4.20±.75
	Q16. Learning to interact with Ryan was easy.	4.00±1.00
	Q17. Talking with Ryan was like talking to a person.	3.90±1.37
	Q18. I felt Ryan understood what I was saying.	4.00±1.26
	Q19. Ryan was friendly.	4.80±.40
	Q20. Ryan was likeable.	4.90±.30
	Q21. Ryan was warm.	4.40±1.28
	Q22. Ryan was intelligent.	4.70±.64
	Q24. Ryan was acting natural.	4.50±.67
	Q25. I would like to interact with Ryan again.	4.60±.66
	Q26. I enjoyed interacting with Ryan at the end of week 3 as much as I did in the beginning of the study.	4.70±.46
	Q27. I found myself looking forward to my sessions with Ryan.	4.70±.64
	Q28. I enjoyed Ryan showing photos to me.	4.90±.30
	Q29. I enjoyed Ryan playing videos for me.	4.90±.30
	Q30. The videos played by Ryan were effective and helped me either learn something new or have a fun conversation.	4.90±.30
	Q31. The conversations were organized and made sense.	4.80±.60
	Q32. If given the change, I would continue further sessions with Ryan.	4.70±.64
Other Questions	Q12. We showed you two versions of Ryan, one with smile and empathy, and one without. Which versions of Ryan do you like the most?	100% selected the version with the smile expression
	Q2/Q9: did you notice a change in the way Ryan communicates with you and her ability in showing facial expressions after the session three crossover?	73% said yes.
	Q33. On the scale of 1-5, how did you like the topics of the conversations Ryan had with you? Response: Kids:4.56, Pets:3.75, TVShows: 3.5 Science:4.50, Music:4.50, Nature:4.88 Foods:4.38, Travel:4.63, Art:3.86, Movies:3.33 Reading:4.13, Sports: 3.44	

C Full survey used in the second AEI study

Table 3: The exit survey evaluating users' perception of Ryan with AEI.

	Question	Avg	Std
1	I felt Ryan was gentle with me.	4.22	1.17
2	Did you notice when Ryan smiled and showed compassion on his face?	Yes (100%)	
3	If answered Yes to Q2, on a scale of 1-5 how would you rate Ryan's facial expressions (Smile and Compassion)?	4.39	0.78
4	I felt happier when I was in the company of Ryan.	3.82	1.13
5	I felt less depressed after talking to Ryan.	3.65	1.27
6	I felt Ryan understood my emotions.	3.28	1.36
7	Ryan encouraged me to open up about my mood and/or feelings.	3.72	1.02
8	The time with Ryan improved my mood and made me feel happier than I was before spending time with him.	3.78	1.11
9	How much do you agree with the statement: "I liked that Ryan empathized with me."	4.00	1.19
10	On a scale of 1-5 how well Ryan did empathize with your feelings?	3.65	1.06
11	Ryan was acting natural.	3.94	1.00
12	Ryan's head/neck movement was natural.	3.94	0.94
13	I liked Ryan's facial expressions.	4.44	0.92
14	I liked Ryan mirroring my facial expressions.	4.00	1.08
15	Ryan's facial expressions were natural.	3.94	1.16
16	The conversation with Ryan was enjoyable.	3.89	1.18
17	The conversation with Ryan was engaging.	3.72	1.18
18	Learning to interact with Ryan was easy.	3.94	1.06
19	Talking with Ryan was like talking to a person.	3.33	1.37
20	I felt Ryan understood what I was saying	3.39	1.09
21	Ryan was friendly.	4.50	0.92
22	Ryan was likable.	4.39	0.98
23	Ryan was warm.	4.06	1.06
24	Ryan was intelligent.	3.83	1.38
25	I felt happier when I was in the company of Ryan.	3.94	1.00
26	Ryan acted naturally.	3.72	1.36
27	I would like to interact with Ryan again.	3.89	1.08
28	I enjoyed interacting with Ryan at the end of the study as much as I enjoyed it at the beginning of the study.	4.06	1.21

Continued on the next page

Table 3 – Continued from the previous page			
	Question	Avg	Std
29	I found myself looking forward to spending time with Ryan.	3.72	0.83
30	I enjoyed Ryan showing photos to me.	3.40	1.26
31	I enjoyed Ryan playing videos for me.	3.80	1.23
32	The videos played by Ryan were effective and helped me either learn something new or have a fun conversation.	3.33	1.50
33	The conversations were organized and made sense.	3.67	0.91
34	If given the chance, I would continue further sessions with Ryan.	3.83	1.04
35	The Ryan calendar is useful for users.	3.81	1.05

D Publications and Patent

- **Hojjat Abdollahi**, Mohammad Mahoor, Rohola Zandie, and Sara H. Qualls. "Artificial Emotional Intelligence in Socially Assistive Robots for Older Adults with Dementia and Depression." *IEEE Transactions on Affective Computing (2022).*

- **Hojjat Abdollahi**, Ali Mollahosseini, Josh T Lane, Mohammad H Mahoor, "A Pilot Study on Using an Intelligent Life-like Robot as a Companion for Elderly Individuals with Dementia and Depression." *IEEE-RAS International Conference on Humanoid Robots*, IEEE, 2017.

- Ali Mollahosseini, **Hojjat Abdollahi**, Timothy Sweeny, Ron Cole and Mohammad H. Mahoor. 'Role of Embodiment and Presence in Human Perception of Robots' Facial Cues." *International Journal of Human-Computer Studies 116 (2018): 25-39.*

- Ali Mollahosseini, **Hojjat Abdollahi**, and Mohammad H. Mahoor. "Studying Effects of Incorporating Automated Affect Perception with Spoken Dialog in Social Robots." *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).* IEEE, 2018.

- Dino, Francesca, Rohola Zandie, **Hojjat Abdollahi**, Sarah Schoeder, and Mohammad H. Mahoor. "Delivering Cognitive Behavioral Therapy Using A Conversational SocialRobot." *International Conference on Intelligent Robots and Systems (IROS)*, 2019.

- Ali Mollahosseini, Behzad Hassani, Michelle J. Salvador, **Hojjat Abdollahi**, David Chan, and Mohammad H. Mahoor. "Facial Expression Recognition from World Wild Web." *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2016.

- Ali Pourramezan Fard, **Hojjat Abdollahi**, and Mohammad Mahoor. "ASMNet: A lightweight deep neural network for face alignment and pose estimation." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1521-1530. 2021.

- Mohammad Mahoor, Josh T. Lane, **Hojjat Abdollahi**, and Eshrat Emamian "Socially Assistive Robot.", US. Patent US-11279041-B2.