

Models of Adaptation in Intelligent Human-Machine Interaction and Their Applications to  
Elder Care and Autism Spectrum Disorder Intervention

By

Jing Fan

Dissertation

Submitted to the Faculty of the  
Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

May, 31, 2019

Nashville, Tennessee

Approved:

Nilanjan Sarkar, Ph.D.

Lorraine Mion, Ph.D.

Gabor Karsai, Ph.D.

Mitchell Wilkes, Ph.D.

Douglas Fisher, Ph.D.

Copyright © 2019 by Jing Fan  
All Rights Reserved

To my amazing parents, Yuejuan An and Xiaozhong Fan, infinitely supportive

and

To my beloved cat, Cuddles, fluffy and affectionate

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Nilanjan Sarkar and Dr. Lorraine Mion for their continued support and help. Their encouragement and help was invaluable and without them this dissertation would not have been possible. I would also like to express my gratitude to Dr. Zachary Warren, Dr. Alexandra Key, and Dr. Paul Newhouse for their continuous help in areas of brain-computer interface, autism spectrum disorder intervention, and elder care. These domain knowledge enabled me to conduct meaningful interdisciplinary research. I would also like to extend my gratitude to other committee members, Dr. Gabor Karsai, Dr. Mitch Wilkes, and Dr. Douglas Fisher, for their willingness and help during my studies and their flexibility.

Second, I would like to acknowledge Dr. Linda Beuscher, Ms. Amy Swanson, and Mr. Gregory Hunt for all of the hard work and continuous help on the user studies in the laboratory as well as in the field. I also appreciate all the graduate and undergraduate students that I have worked with during my dissertation research. Specially, thanks go to Joshua Wade, Esube Bekele, Yi Li, Sandeep Srinivasan, Akshith Ullal, and Dayi Bian. Furthermore, I want to acknowledge other wonderful graduate students of the Robotics and Autonomous Systems Lab, Lian Zhang, Zhi Zheng, Huan Zhao, Guangtao Nie, Zhaobo Zheng, and many others for frequent discussions, feedback, and support. I also would like to thank all my friends at Vanderbilt University for our enjoyable friendship.

Finally, I would like to express my heartfelt gratitude to my family. My parents, Yuejuan An and Xiaozhong Fan, have provided me love, support, and encouragement to help me throughout my entire life. My cat, Cuddles, has accompanied me the most during this long journey.

This dissertation work have been supported in part by the National Institute of Health under Grant 1R21AG050483-01A1 and Grant 1R01MH091102-01A1, the National Science Foundation under Grant 0967170, the Vanderbilt Kennedy Center through the Hobbs Society Grant, and the Vanderbilt University School of Nursing Independence Foundation Chair Funds.

# TABLE OF CONTENTS

	Page
DEDICATION .....	iii
ACKNOWLEDGEMENTS .....	iv
LIST OF TABLES .....	ix
LIST OF FIGURES .....	xi
Chapter	
1 Introduction.....	1
1.1 Background.....	1
1.1.1 Elder care and technology.....	1
1.1.2 ASD intervention and technology.....	3
1.2 Scope of this Work.....	4
1.3 Literature Review.....	6
1.3.1 SAR systems for older adults.....	7
1.3.2 SAR systems for individuals with ASD.....	12
1.3.3 Intelligent HCI systems for people with special needs .....	16
1.3.4 Summary .....	17
1.4 Structure of this Dissertation.....	18
REFERENCES .....	20
2 EEG-based Affect and Workload Recognition.....	28
2.1 Brief Summary.....	28
2.2 Abstract.....	30
2.3 Introduction.....	30
2.4 System Description and Data Acquisition .....	32

2.4.1 Virtual driving environment.....	33
2.4.2 Data acquisition .....	34
2.4.3 Subjects .....	35
2.5 Methods.....	35
2.5.1 Signal processing .....	36
2.5.2 Feature generation.....	36
2.5.3 Feature calibration.....	37
2.5.4 Feature selection and classification.....	38
2.6 Results.....	39
2.6.1 Classification results .....	39
2.6.2 Discriminative features .....	41
2.7 Discussion.....	44
2.7.1 Validation of therapist’s measures of mental states.....	44
2.7.2 Comparison with related works .....	46
2.7.3 Limitations and future works .....	47
2.8 Conclusion .....	48
REFERENCES .....	49
3 A Robotic Coach Architecture (ROCARE).....	52
3.1 Brief Summary.....	52
3.2 Abstract.....	53
3.3 Introduction.....	54
3.4 Design of a Robotic Coach System Architecture.....	56
3.4.1 Comparison with existing SAR architectures .....	56
3.4.2 Engagement models .....	58
3.4.3 Sensing and actuation.....	59
3.4.4 Database and graphical user interface.....	60
3.4.5 Supervisory controller.....	61
3.5 Feasibility Studies .....	62
3.5.1 HRI scenarios.....	62
3.5.2 Participants and protocol.....	65
3.6 Results.....	66
3.6.1 Data analysis methods.....	66
3.6.2 One-on-one interaction results .....	67

3.6.3 Triadic interaction results.....	68
3.7 Discussion.....	70
3.8 Conclusion.....	71
REFERENCES.....	73
4 Robot-Mediated Social Interaction within a Multimodal Task.....	76
4.1 Brief Summary.....	76
4.2 Abstract.....	77
4.3 Introduction.....	78
4.4 Challenges for Robot-mediated Social Interaction.....	80
4.5 HRI Framework.....	81
4.6 System Design and Development.....	84
4.6.1 VR-based multimodal task.....	84
4.6.2 Supervisory controller.....	87
4.6.3 Objective measures.....	91
4.7 User Study.....	94
4.7.1 Experimental design.....	94
4.7.2 Participants.....	95
4.8 Results.....	96
4.8.1 System performance results.....	96
4.8.2 Objective measures of interaction.....	97
4.9 Discussion and Conclusion.....	101
REFERENCES.....	104
5 Field Testing of Robot-Mediated Triadic Interaction.....	106
5.1 Brief Summary.....	106
5.2 Abstract.....	107
5.3 Introduction.....	107
5.4 Method.....	110
5.4.1 Robotic system.....	110
5.4.2 Field study.....	113
5.4.3 Data collection and analysis.....	115
5.5 Results.....	119

5.5.1 Objective data analysis results .....	119
5.5.2 Subjective data analysis results .....	122
5.6 Discussion and Conclusion .....	124
REFERENCES .....	126
6 Mathematical Model of Adaptation for Multi-User HRI .....	130
6.1 Brief Summary .....	130
6.2 Abstract .....	130
6.3 Introduction .....	131
6.4 Computational Framework .....	134
6.4.1 Problem formulation .....	134
6.4.2 Formal modeling .....	136
6.5 Simulation .....	139
6.5.1 Multi-user HRI scenario .....	139
6.5.2 The MOMDP model of <i>RockSample</i> .....	140
6.5.3 User simulation .....	142
6.5.4 Results .....	144
6.6 Conclusion .....	146
REFERENCES .....	148
7 Contributions and Future Work .....	151
7.1 Overall Contributions .....	151
7.2 Technical Contributions .....	152
7.2.1 Models of people, interaction, and machine .....	152
7.2.2 SAR systems development .....	154
7.2.3 Automatic evaluation methods .....	155
7.3 Contributions to the Science of Elder Care and ASD Intervention .....	156
7.4 Future Work .....	157



## LIST OF TABLES

Table	Page
2-1. Subject Characteristics .....	35
2-2. Extracted Features .....	37
2-3. Classification Results Using the Best Performing Features and Hyper-parameters.....	41
2-4. Error Counts .....	46
3-1. Survey Results for One-on-one Experiment N=11.....	67
3-2. Survey Results for Triadic Experiment N=14 .....	69
3-3. Summarized EEG Engagement Index for Six Participants .....	69
4-1. Variables Related to Feedback States.....	91
4-2. Effort Analysis Results.....	98
4-3. Post-test Results.....	99
4-4. Head Pose Analysis Results.....	100
4-5. Vocal Sound Analysis Results.....	101
4-6. EEG Analysis Results.....	101
5-1. Experimental Protocol .....	114
5-2. Validation Results of Automatic Evaluation Algorithms .....	117

5-3. Peripheral Physiological Features.....	118
5-4. Participants' Engagement across Six Sessions .....	119
5-5. Classification Results .....	122
5-6. RAS Results.....	122
5-7. VAS Results .....	123
5-8. Post Experiment Evaluation Results.....	123

## LIST OF FIGURES

Figure	Page
1-1. A Visual Summary of Elements Involved in HMI for People with Special Needs .....	5
2-1. Framework Overview .....	33
2-2. Comparison of 2D Feature Scatter Plot for Engagement (a) Before Feature Calibration, (b) Baseline Feature Distribution, (c) After Baseline Removal, and (d) After Feature Normalization .....	38
2-3. Classification Results Evaluated Based on Macro-averaged $F_1$ Score .....	40
2-4. Feature Usage and Electrode Usage .....	43
2-5. Overall Rating Data as a Function of Participants' Performance .....	45
2-6. Learning Effect .....	45
3-1. Robotic Coach System Architecture (ROCARE) .....	57
3-2. Timed Automaton Model of $ES_a^i(t)$ .....	59
3-3. (a) Supervisory Controller Module for One-on-One Interaction (b) Control Policy for Triadic Interaction .....	65
3-4. Summarized EEI (left) and Summarized GAS (right) as a Function of Self-rated Activity Preferences .....	68
3-5. Participants' Head Pose Yaw Angles during Triadic Experiment .....	70
4-1. Task Structure .....	82
4-2. System Architecture .....	83

4-3. Virtual Environment .....	85
4-4. Interaction Boxes .....	85
4-5. Supervisory Controller Module .....	88
4-6. Raw Sound Source Angle and Yaw Angle Data for One Pair.....	93
4-7. Experimental Setup and Procedure.....	95
4-8. Effort Analysis Results .....	98
5-1. Ro-Tri Architecture .....	111
5-2. Ro-Tri System Overview and Setup at a Retirement Community.....	113
5-3. Raw Sound Source Angle and Head Pose Yaw Angle Data for One Session.....	116
5-4. Changes of Engagement Over Six Sessions .....	121
6-1. Examples of Interaction Flows among the Human-Robot Team.....	134
6-2. Example <i>RockSample</i> Task Environment.....	139
6-3. Case 1 Simulation Results .....	144
6-4. Case 2 Simulation Results .....	145
6-5. Case 3 Simulation Results.....	146

# CHAPTER 1

## INTRODUCTION

### 1.1 Background

Human-machine interaction (HMI) is a field that studies the interaction and communication between user(s) and machine(s) via a human-machine interface. The whole system, which is called human-machine system (HMS), consists of at least three main components: the user, the machine, and the interaction between the user and the machine [1]. The machine in a HMS can be a robotic platform, a computer, a mobile device, an embedded system, and so forth. There are two general categories of interaction, which are remote interaction and proximate interaction [2]. In remote interaction, the user and the machine are separated spatially or even temporally. Whereas in proximate interaction, the user and machine are collocated. Given the differences in the needs of the users, the properties of the machines, and the types of the interaction, there are vast potential for applications of HMI, ranging from teleoperation to service robots in human-robot interaction (HRI) [2] and from remote collaboration to driver monitoring in human-computer interaction (HCI) [3]. The motivation for this work stems from the increasing demand for innovative technologies and HMSs to facilitate and augment the delivery of effective and safe care for older adults as well as to provide treatments for individuals with autism spectrum disorder (ASD).

#### *1.1.1 Elder care and technology*

The population in the US is aging rapidly as the first batch of baby boomers started turning 65 in 2010. The number of older people is projected to represent nearly 21 percent of the total population in 2030, which is twice as large as the number in 2000 [4]. With aging, many older adults experience chronic health conditions, functional limitations, dementia, and problems with physical functioning, falls, and mental health [4-7]. Dementia, including Alzheimer's disease and other related disorders, is overwhelmingly faced by older adults. One in ten people age 65 and over has Alzheimer's disease in the USA. Dementia impacts communication and interaction ability, impairs judgement, memory, and affect regulation. An additional 15 to 20 percent of older adults have mild cognitive impairment (MCI) and are at higher risk of later developing dementia [8]. The health care costs for older adults with concomitant medical conditions and physical and cognitive impairments are substantial [9, 10]. Informal unpaid caregivers such as family and friends provide 83 percent of the assistance and are under high financial, emotional, and physical burden

[10, 11]. Thus, there is an urgent need for technological strategies that can coexist within resource strained environments to augment the process of effective care for older adults.

There are five general categories of dementia healthcare technologies [12], which are i) diagnosis and assessment, such as neurocognitive testing via video telemedicine [13] and sensor-based early sign and progression detection [14, 15]; ii) monitoring, using wearable sensors and distributed sensor networks to monitor older adults' health, behaviors, and activity [16]; iii) assistive, such as wandering prevention tools, cognitive orthotics that provide activity reminder and medication management [16], and intelligent systems that can facilitate activities of daily living (ADL) [17]; iv) therapeutic, mostly robotic systems such as animal robots to provide companionship and address mental illness [18], telepresence robots to facilitate social connections with families and caregivers [19], and socially assistive robotic (SAR) systems to provide activity-oriented therapies such as physical exercise and memory games [20]; and v) caregiver-supportive, such as multimedia systems to facilitate communication between older adults with dementia and caregivers [21]. The majority of the HMSs for older adults are distributed systems (ambient assisted living systems in particular) and personal service robots for the purpose of supporting older adults in the completion of ADLs and monitoring their behaviors and their environment in order to enhance independent living [22]. Technological advancements in SAR and virtual reality, together with a growing body of literature on risk reduction and prevention and non-pharmacologic therapies, have gained momentum in recent years for therapeutic technology.

Evidence suggest that exercise and physical activity, lifelong learning/cognitive training, and healthy diet may reduce the risk of cognitive decline and dementia. Evidence is growing that social isolation is a risk factor for dementia and social and cognitive engagement may reduce such risks [8, 12, 23]. Physical, cognitive, and social activities have also been shown to improve older adults' physical and psychological well-being and reduce the risks of many health problems such as falls [4, 24, 25]. Although neither pharmacologic nor non-pharmacologic therapies can treat dementia or slow or stop their progression at present, reviews and meta-analyses indicated that cognitive intervention, exercise and physical activity intervention are beneficial to people with Alzheimer's disease and have positive effects on cognitive function [26-28]. Instead of favoring a single intervention, the literature on non-pharmacologic therapies suggests multimodal strategies that tailored to the individual and highlights the importance of social engagement in addition to older adults' physical and mental health [12, 29, 30]. It is within this context, I propose models of adaptation and HMSs to deliver multimodal therapies with an emphasis on social engagement.

### **1.1.2 ASD intervention and technology**

ASD is a prevalent and fast-growing developmental disability characterized by social communication impairments and restricted, repetitive patterns of behavior [31]. An estimated 1 in 59 children in the US has been diagnosed with ASD [32]. At present, there are no medications to cure ASD or treat the core symptoms. Previous research has established that early intensive intervention, behavior and communication approaches, are efficacious and can improve a child's development [33, 34]. A small but growing body of literature has investigated behavioral and educational intervention on meaningful skills related to adaptive adult independence in order to support the transition of youth with ASD to adulthood [35]. However, the financial cost for these interventions are high. Intensive behavioral interventions for children with ASD cost \$40,000 to \$60,000 per year for each child [36]. Hence, innovative technologies that could target specific core symptoms and meaningful skills for individuals with ASD are in urgent need.

Individuals with ASD show an affinity towards technologies, which leads to the use of computer programs, virtual reality, and robotics to achieve specific therapeutic objectives [37-39]. Technological systems have the advantages of providing immediate, predictable, and repeated responses within a safe and non-threatening environment [39]. This allows individuals with ASD to focus their attention on practicing repeatedly a specific skill. In the ASD intervention literature, researchers have proposed systems to different age groups ranging from toddlers to adolescents. In HRI, robotic systems have been used to explore the response of individuals with ASD to robots in comparison to human, to elicit target behaviors, to teach and practice skills such as imitation and joint attention, and to reinforce skill learning through feedback and encouragement [40]. In HCI, computer programs and virtual reality environments have been designed to train skills including social problem solving, facial and emotional processing, spatial planning, language skills, academics and cognitive skills, and skills for independent living such as driving [38, 39, 41].

Many HMSs for ASD intervention are structured in the form of one-to-one HMI, in which a single individual with ASD interacts with a single robot, computer program, or virtual reality environment. In recent years, there is an emergence of many-to-one HMI for ASD intervention. These include using a robot mediator to facilitate social exchange between a child with ASD and a partner [42, 43], developing collaborative virtual environment with distributed architecture to support remote collaborative interaction between children with ASD [44, 45], and designing collaborative tasks for two or more children with ASD on touch-sensitive shared active surfaces that can be operated by multiple people simultaneously [46, 47]. These studies indicate that a robot mediator could facilitate human-human interaction and collaborative interaction has positive impacts on social skills of individuals with ASD.

In the following sections of this chapter, I first establish the scope of my research in Section 1.2. Section 1.3 presents the past research as well as current state-of-the-art on my research topics, including a detailed survey and discussion of SAR systems for older adults in Section 1.3.1, a detailed survey and discussion of SAR systems for individuals with ASD in Section 1.3.2, and significant works on intelligent HCI systems for older adults or individuals with ASD in Section 1.3.3. Finally, Section 1.4 describes the structure of this dissertation.

## **1.2 Scope of this Work**

Figure 1-1 summarizes the elements involved in HMI for older adults and individuals with ASD, who I referred to as people with special needs in the following paragraphs. In this work, I mainly focus on proximate HMI with SAR systems. Virtual reality environment and computer programs are used as tasks and to elicit human mental states related to models of people. With respect to this dissertation, I would like to create models for people with special needs. In a typical HMI scenario, one or multiple human users are either involved in free-form interaction or guided through predetermined task or behaviors with the robot taking different roles including a subordinate team member, facilitator, coach, and peer. In addition to HMI, when multiple people interact with SAR, researchers often observe human-human interaction that are generated due to HMI. In these SAR studies, usually an administrator is present to remotely operate the robot, observe human behaviors, monitor or intervene the interaction between people with special needs and the robot. There are three types of model in HMI for people with special needs. Models of people are used to understand human actions, behaviors, and their mental states. Models of interaction create mechanisms to guide HMI or even HHI given the context of the task as well as information from models of people and models of machines. Models of machines capture the machines' ability to learn their interaction with human, such as how their behaviors affect human interaction.



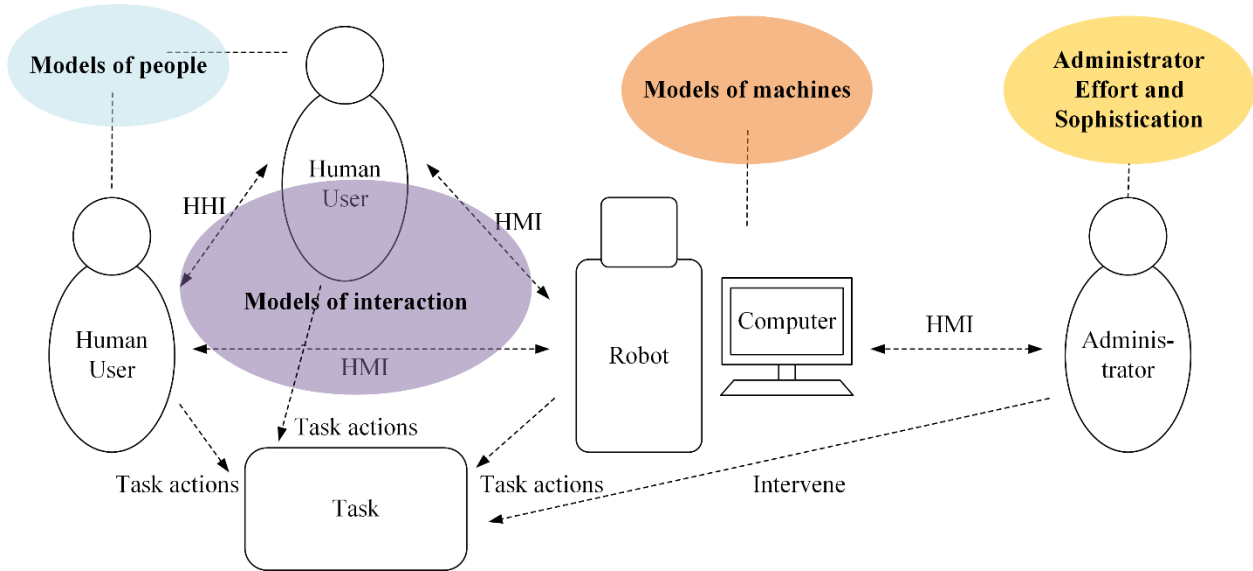


Figure 1-1. A Visual Summary of Elements Involved in HMI for People with Special Needs

Many SAR systems that are being developed and evaluated for people with special needs are either controlled remotely using the Wizard of Oz (WoZ) experimental paradigm or restricted to pre-programmed robotic behaviors. The behaviors of the SAR system with WoZ design are perceived as adaptive. However, the administrator is responsible for constantly evaluating human behaviors, human interaction with the system, and controlling the SAR system. Therefore, the adaptation in WoZ design is achieved at the cost of great administrator effort and sophistication. Other SAR systems that rely solely on a sequence of pre-programmed robotic behaviors are called open-loop systems and are not adaptive. These systems are limited in their capacity for HRI and often require carefully designed tasks to reduce user frustration and administrator intervention. For these two types of SAR systems, there are no models of people, interaction, or machines.

SAR systems that are able to react to user behaviors and the context of the interaction are closed-loop systems. In recent years, there has been an increasing amount of literature on closed-loop SAR systems for people with special needs. These systems are integrated with sensing modules to automatically detect user behaviors such as their gesture performance [48-50], proximity with respect to the robot [51], focus of attention [52], body postures [20], tactile interaction [53], and task actions [20]. Human behaviors encompasses both explicit human behavior and the implicit mental states hidden behind behavioral performance. A few SAR systems have developed models of people to understand mental states including attentive or distracted behavior based on body language [20], affective and cognitive states based on physiological signals [54] and eye gaze data [55], and positive or negative interaction based on physical

locations [51]. One goal of this work is to further investigate mental state models of people including affective and cognitive states as well as intention, which represents context-sensitive mental states.

The models of interaction in existing closed-loop SAR systems have two limitations. First, most systems to date have predominantly focused on one-to-one interaction. Multi-user interaction is pivotal for fostering social interaction for people with special needs. Second, almost all the previous work were designed within the scope of a particular interaction scenario to engage older adults in a physical or cognitive activity or to teach individual with ASD a single skill. As a consequence, the models of interaction are task-specific and cannot be reused for a different interaction scenario. Therefore, the goals of this work with respect to models of interaction is to design a system architecture that is not bound to a specific task and to develop SAR systems to support many-to-one interaction in order to foster human-human interaction through HMI.

Although researchers have designed and developed different behaviors of robots used in SAR systems, including their motion, gaze behavior, emotional behavior, and facial expressions, to my knowledge, there is very limited work on models of machine that enable the robot to understand how its behaviors affect human behaviors. An adaptive system not only needs to respond autonomously to human behaviors, but also required to adapt based on human responses to robot behaviors. Another goal of this work is to design such models of machines.

To summarize, my research focuses on the models of people, interaction, and machines in order to design and develop intelligent HMSs for people with special needs. This work pushes the boundaries of the adaptive automation capabilities of SAR systems for people with special needs, and offers SAR systems to engage older adults in multimodal activity-oriented therapies in the form of both one-to-one interaction and triadic interaction to enforce human-human interaction.

### **1.3 Literature Review**

The main focus of this work is proximate HMI with SAR systems. In this section, SAR systems designed and developed for people with special needs are reviewed in detail. Intelligent HCI systems that are relevant to the scope of this work are also reviewed. The term SAR is defined by Feil-Seifer and Mataric in 2005 as robotic systems that provide assistance to human through social interaction rather than physical contact [56]. SAR belongs to the intersection of assistive robotics and socially interactive robotics in that the goal of the robot is to give assistance and achieve measurable progress as in assistive robotics and such goal is achieved by creating close and effective social interaction as in socially interactive robotics. SAR systems are motivated to address user populations such as older adults, post-stroke patients, individuals

with ASD, and students where social interaction rather than contact is the central focus of the designated assistive tasks. The application of SAR for older adults and individuals with ASD is also motivated by the fact that safety risk decreases significantly due to non-contact HRI.

### **1.3.1 SAR systems for older adults**

#### *1.3.1.1 Robotic platform and interaction scenario*

Research on SAR for older adults is being conducted in several countries to provide support for independent living, to monitor health and safety, to provide companionship, and to provide activity-oriented therapies. The physical appearance of robots can be broadly categorized into three groups: machine-like appearance, animal-like appearance, and human-like appearance. Machine-like robotic platforms are mostly used as service robots that facilitate independent living by providing a range of services. Many of the service robots are integrated with a touch screen control and some of these robots also monitor older adults' health and safety. Coradeschi et al. developed GiraffPlus system with the Giraff telepresence robot to continuously monitor activities of older adults, to provide warnings, alarms, and reminders, and to encourage social contact through the Giraff robot [19]. The Giraff robot is a human-height mobile robot with a LCD panel, a camera, a speaker, and a microphone for video conferencing [57]. Gross et al. developed a home robot companion that has a mobile base, a touch-screen, and two eye displays [58]. The robot has a set of functionalities such as navigation, search user behavior, active daytime management, encouragements to do cognitive training, and making video calls to family members.

Animal-like robotic platforms are primarily used to provide companionship similar to animal-assisted therapy. The goal of these robots is to improve older adults' relaxation and motivation, psychological wellbeing, and provide social support. Researchers have developed a robotic seal, Paro [59], a robotic dog, AIBO [60], and a robotic cat, NeCoRo [61], that can respond to visual, aural, and tactile stimuli and behave like an animal. Interacting with animal robots have shown positive effects on older adults including increase in engagement activities, decrease in psychological stress reactions, increase in pleasure and interest, improvement in speech and emotional words, increase in quality of life score and decrease in loneliness [62-65]. Animal-like robotic platforms are also used for health-monitoring and providing information support. Hopis is a dog-like fluffy robot that is able to take blood pressure, body temperature, and question the person about their health [66]. Nabaztag [67], a rabbit-like robot, and iCat [68], a cat-like robot, were used to encourage older adults to maintain a healthy lifestyle through conversation.

Human-like robotic platforms are the most comprehensively designed and widely used for older adults. Pollack et al. developed a mobile robotic assistant Pearl to provide older adults with reminders about their daily activities and to help older adults navigate their environments [69]. Like other service type robots, Pearl has a mobile base and a touch sensitive graphical display. The human feature of Pearl is an actuated head units capable of facial expressions. Graf et al. developed a robotic home assistant Care-O-bot II, which contains a mobile platform with a touch screen and is equipped with adjustable walking supporters and a manipulator arm [70]. Care-O-bot II has a head with static face and was developed to provide physical assistance including walking aid and fetch and carry task. Khosla et al. developed a small human-like robot with a baby face appearance, Matilda, to provide services and companionship [71, 72]. Matilda, which belongs to NEC's PaPeRo family of robots, does not have a mouth, arms, or legs. It is integrated with smart phone, touch panel, and remote computer to carry out services including singing and dancing, playing cognitive games, placing video call, reminding, weather forecasting, walking and delivering exercise dialogue, and providing diet suggestion. Inoue et al. used PaPeRo to keep older adults informed of their daily schedule and prompt them to take desired actions [73].

In addition to providing services, human-like robots have been used to facilitate therapeutic intervention through robot-led activity-oriented therapies. Bandit [48], NAO [74, 75], Manoi-PF01 [76], RoboPhilo [77], AprilPoco [78], and TAIZO [79] were used to encourage and instruct older adults to perform physical exercises. NAO, Manoi-PF01, RoboPhilo, and TAIZO are biped humanoid robots that can generate whole body motions. AprilPoco is an 11-inch tall robot made by Toshiba with movable head, two arms, and a base that can rotate. Bandit was developed by mounting a humanoid torso on a Pioneer 2DX mobile robotic platform. Tapus et al. tested the effectiveness of robot-mediated cognitive intervention (song discovery game) using Bandit and observed an improvement in task performance for three older adults with dementia [80]. McColl et al. developed humanoid robot Brian 2.1 to engage older adults in eating activity and a cognitive stimulation activity (memory card game) [20]. Engagement activities to reduce depression and solitude of older adults in health care facilities was conducted using YORISOI Ifbot, including quizzes, riddles, music, simple math, and tongue twisters [81]. Louie et al. developed a humanoid Tangy with a head and two arms to provide group-based cognitive intervention (Bingo game) [82].

Different aspects of a robot's appearance and behaviors affect the human users in a variety of ways. Robotic platforms developed to provide support for independent living are usually human-height mobile robot with a touch screen. Some of these robots are integrated with a cartoon-like or human-like head for the purpose of expressing emotional states. Robotic platforms developed for emotional wellbeing of older adults and to provide companionship are usually small robots with animal shape or baby-like appearance. In the context of providing activity-oriented therapies, humanoid robots are more powerful given their

unique ability to demonstrate physical exercises or play cognitive games with older adults. Besides, they are able to provide much more human-like social cues, such as facial expressions, gaze behaviors, gestures, and body language. These social factors are likely to increase the social presence of the robot, enhance user's motivation, and further improve user's task compliance [68, 83].

### *1.3.1.2 Modes of human-robot interaction*

Speech has been the preferred natural means of communication in SAR. It is easy to program a robot to speak a language. However it has been challenging to develop speech recognition software to understand the variation in human utterances. Thus most robot speech-based communications have been either one-way (i.e., the robot speaking only) or have very rudimentary and restricted capabilities of understanding of spoken words. In addition to verbal communication, SAR systems are embedded with cameras and sensors to allow for various modalities of non-verbal communication between a robot and an older adult. Laser range finder and sonar sensors are used for obstacle detection, navigation, and person tracking [58]. Cameras and RGB-D sensors such as Kinect together with computer vision technique allow older adults to interact with SAR through movement, gesture, posture, and facial expression. Robot Tangy recognized older adults' request for assistance by a raised hand gesture [82]. Robot Brian 2.1 inferred states of older adults, whether they were attentive or distracted, based on their trunk orientation and face orientation [20]. Robot Bandit monitored the arm movements of older adults in order to evaluate their performance of chair exercises [48]. Tactile sensing is another modality for non-verbal communication. Bhuvaneswari et al. used the tactile sensor on NAO's head for users to give acknowledgements [84]. Animal robots Paro, AIBO, and NeCoRo have tactile sensors on their heads, chins, and bodies to register petting by older adults. Key inputs, buttons, and remote controllers have also been integrated with SAR as means for interaction. Fasola and Mataric used a Nintendo Wiimote wireless Bluetooth button control interface with three buttons to allow older adults to communicate with the robot [48]. In [67], older adults communicate with the Nabaztag by pressing yes or no conversation buttons. This method is simple to realize and usually has a higher input accuracy, but its acceptance has not been extensively tested with older adults with cognitive impairment. Lastly, touch screen and LCD display are widely used by service robots.

In human-human interaction, gestures, postures, gaze, face orientation, and facial expressions are effective ways to facilitate the delivery of communicative contents, and to exhibit one's emotional state, intention, and openness to conversation. Likewise, SAR systems are developed to express similar social cues for the purpose of interacting with older adults in a socially appropriate manner. Gaze, head movement, and gestures such as pointing, hand waving, and celebration gesture are widely used in the existing SAR systems [20, 72, 85]. Despite the vast application of these robot behaviors, researchers generally do not

study the effect of different robot body language on the social presence and trustworthiness of SAR. More attention is given to the expressiveness of emotional states. Robot Brian 2.1 has a 5-DoF facial muscle system to display emotions such as happy, neutral, and sad [20]. Robot Bandit contains 1-DoF expressive eyebrows and 2-DoF expressive mouth [48]. Animal-like robot iCat expresses emotions such as happy, sad, and understanding by moving its lips, eyebrows, eyes, eyelids, head and body [68]. Zecca et al. developed a SAR robot KOBIAN that is capable of whole body emotion expressions by combining facial expression and body posture [86]. Without an expressive face, emotional expression behaviors are generated by programming different patterns and speeds of robot movements and colors [87, 88]. In addition to displaying emotional states through visual communication channel, aural communication channel is also utilized by varying robot's speaking rate and vocal pitch [20].

HRI is also characterized by the structure of the interaction trials between older adults and robots. Most systems to date have predominantly focused on one-to-one interaction. Many-to-one interaction is pivotal for fostering social interaction. Louie et al. developed an autonomous assistive robot Tangy that plays Bingo game with a group of older adults [82]. However, the goal of the system is to plan, schedule, and facilitate group activity instead of promoting interpersonal social interaction. Tangy is responsible for leading the Bingo game and facilitate any individual older adult one at a time. Similar to Tangy, robot Matilda was designed to play Bingo and Hoy with groups of 8 to 30 older adults [71]. Matilda communicates with older adults using voice commands or a remote touch panel. No robot behaviors to facilitate interpersonal social interaction were reported. Back et al. [74] and Matsusaka et al. [79] developed SAR systems to lead physical exercise with multiple older adults. Konah et al. [81] developed a series of robot assisted activities for group interaction. These systems have been shown to be useful, however, they either operate in an open loop fashion or require a human mediator, and thus are limited in their ability to facilitate social interaction among older adults. Although not designed directly for older adults, one SAR system was developed to enhance human-human interaction. Matsuyama et al. programmed a conversational robot SCHEMA to participate in a conversation game with the goal of promoting the communication activeness of human participants [89].

### *1.3.1.3 Models of adaptation*

The adaptive automation capabilities of SAR systems are captured by models of people, interaction, and robots as described in Section 1.2. The most common scheme of adaptive automation is a combination of models of explicit human behaviors and a rule-based, task-specific model of interaction given a library of possible robot behaviors. Gorer et al. [75] used a Kinect sensor to track the skeleton of an older adult and applied the dynamic time warping method to compare older adult's exercise performance with the

stored exercise template. The robot then provided feedbacks based on predefined rules and the detected user exercise performance. Fasola et al. [48] recognized user's arm poses by detecting user's face, hand, arm locations, and arm angles from image sequences taken by a camera. The robot behavior was governed by a finite state machine that took into account task performance, progress, session history, and added variability to the same verbal feedback by introducing filler words in order to personalize HRI and maintain user engagement. Robot Brian 2.1 [20] determined user's meal-eating progress by information from the meal-tray-sensing platform that tracked weight changes of meal items and the utensil-tracking system that evaluated relative position of utensil with respect to the user's head and the direction of motion of the utensil. A set of robot behaviors was generated based on a finite state machine to engage older adults in eating activity with mechanisms to incorporate robot's emotional states in accordance to its verbal feedback. Many other SAR systems have models of people to understand explicit human behaviors including simple speech, gesture, tactile interaction, and task-related action [59, 76, 77, 80, 82, 90].

In contrast to SAR for activity-oriented therapies, service type SAR systems usually have a more generalized models of interaction due to the fact that these robotic systems were designed with the aim to provide a range of services. In general, service type SAR systems are designed following a layered system architecture [71, 72, 91]. The lowest layer corresponds to the basic sensors and actuators on the robotic platform. The highest layer corresponds to different types of services that the robot needs to provide. In the middle, there is a layer for specific primitive functionality such as face recognition and localization, a layer for more complex functionality such as follow a user and navigation, and a layer that combines functionality and task such as a dialog manager. This type of models of interaction can be considered as a functionality-centric model. However, such models are not directly applicable to SAR systems for activity-oriented therapies because a more user-centric model is needed to engage older adults in activity-oriented therapies, which is not the main focus for a layered system architecture.

Only a few SAR systems were integrated with models of implicit mental states. Khosla et al. [72] detected user's emotional response based on the changes in facial action units and relied on history of emotional responses to different services to determine older adult's preference and personalized its service according to individual preferences. In another work by Khosla and Chu [71], the dialogue between Matilda and an older adult was estimated to be in one of the five mental states related to the model of behavior change, which were pre-contemplation, contemplation, preparation, action, and maintenance. Matilda then adapted its behavior based on the mental state estimation, user emotion, and speech. Robot Brian was developed to estimate attentive or distracted user state based on trunk and face orientation [20], affective states including stressed, bored, neutral, and positively excited based on verbal intonation [92], and affective arousal level based on heart rate [93]. To the best of my knowledge, only one work adapted the

robot behavior based on human responses to robot behaviors. Chan et al. [93] applied MAXQ learning to allow a robot to learn its interaction with human and take personalized actions to improve a user's state from the stressed state to the non-stressed state while engaging the user in a cognitively stimulating activity.

### **1.3.2 SAR systems for individuals with ASD**

The first application of robot for individuals with ASD dates back to 1976, when Emanuel observed verbal and nonverbal communication and social interaction of one child with ASD during play time with a LOGO turtle robot [94]. Over the past four decades, SAR systems have been used to explore the response of individuals with ASD to robots in comparison to human, to elicit target behaviors, to teach and practice skills such as imitation and joint attention, and to reinforce skill learning through feedback and encouragement [40]. There are many similarities between SAR systems developed for older adults and those developed for individuals with ASD. First, the physical appearance of robots are either machine-like, animal-like, or human-like. Some robotic platforms such as YORISOI Ifbot, NAO, and Bandit were used for both populations. NAO humanoid robot is the most widely used robotic platform for ASD intervention. Second, the robots interact with human through verbal and nonverbal communication. Depending on the physical appearance and capabilities of the robotic platform, nonverbal communication such as facial expression, body language, movement pattern, tactile interaction, and gaze behavior were considered as part of the HRI design. To avoid duplications of similar information, in this section, I focus on the design of HRI and the applied models of adaptation.

#### *1.3.2.1 Design of human-robot interaction*

There are many works in the ASD literature that compared robot-mediated therapies with human-mediated therapies. Recent literature review [95] summarized observations from robot-mediated therapies, including: i) individuals with ASD performed the task better in robot condition compared to human condition; ii) in some cases, individuals with ASD responded to robots rather than to human, which is the opposite for their typically developed (TD) peers; and iii) higher levels of stimulation were better than lower levels of stimulation. These results are encouraging towards the application of SAR for ASD intervention.

The ultimate goal of SAR systems for individuals with ASD is to improve their social skills, emotional awareness, and communication skills, and this goal is pursued by designing HRI to elicit target behaviors in robot-mediated therapies. Researchers have attempted to develop robotic platforms and design HRI for a number of behaviors and skills.



*Imitation* Imitation plays a role in the transfer of knowledge from an external source to the individual with ASD. Greczek et al. [50] used NAO to play an imitation game called “Copy-Cat” where the robot demonstrated one of ten arm poses and asked the child to imitate. The child received sensory rewards and encouragement from the robot based on one’s imitation performance. If the child failed to imitate the robot pose, the robot would gradually intensify its stimulation by choosing a higher level of prompts. Zheng et al. [96, 97] used NAO to play a two-way imitation game with children with ASD. NAO first imitated the child and then asked the child to imitate one of its four gestures or to imitate a sequence of gestures. Duquette et al. [98] developed a human-like robot Tito, a 28 inches tall robot with two arms, a rotatable head, a mobile base, and a static face that can express smile, to engage children in a series of imitation play patterns. There were three levels of imitation, which were facial expression, body movements, and familiar actions. In general, a robot teaches this skill by engaging the individuals with ASD in an imitation game and reinforce skill learning through feedback and encouragement.

*Joint Attention* Joint attention is the act of sharing attention with others such as pointing, showing objects, and coordinating gaze. Anzalone et al. [99] used NAO to induce joint attention with children by gazing, by gazing and pointing, and by gazing, pointing, and vocalizing at pictures. The robot alternated its gaze toward the child and then toward the picture and gradually intensified its stimulation by adding gesture and gesture and speech. Kozima et al. [100] developed Keepon, a small animal-like yellow robot with only four motors on the whole body, to interact with children. Keepon has two types of interactive actions: attentive action by directing its head left/right and up/down, and emotive action such as pleasure and excitement by rocking left to right and bobbing up and down. In their study, Keepon alternated its gaze between a child, a caregiver, and a nearby toy and produced emotive actions upon any meaningful child actions. Bekele et al. [101] and Zheng et al. [102] used a NAO robot to prompt a child to look at left or right target monitors. There were six levels of prompts following the “least-to-most” hierarchical protocol. In general, a robot teaches this skill by guiding the child’s attention to a specific object and providing feedback as the child is making progress.

*Social Behavior* Social behavior is the most widely investigated behavior in the ASD literature. Previous studies have explored the application of SAR to address social skills including eye contact, turn-taking, self-initiated interaction, and emotion recognition and expression. Shamsuddin et al. [103] programmed NAO to carry out HRI modules composed of simple robot behaviors such as eye blinking, head-turn, moving arms, and playing music. The SAR was developed to target communication behavior with the aim to engage children to communicate with the robot. Costa et al. [104] used a LEGO MindStorms NTX to play a simple game with children with ASD. The child and the robot took turns to kick a ball to each other. When another child was involved, the robot was controlled by one child to play the game with the other

child in order to establish social behavior between the two children through the robot mediator. Mazzei et al. [105] developed FACE, consisting of a passive body with an active female head that encapsulated 32 motors to stimulate and modulate human facial expressions, in order to teach individuals with ASD emotion recognition, facial expression, and empathy. Damm et al. [106] used a robotic head Flobi with human head and face features to study gaze behaviors of individuals with ASD in HRI. Flobi conveyed its preference to one of the two cards by directing its eye gaze. The results showed that individuals with ASD had significantly more eye contact with the robot than with a human actor. Robins et al. [42] developed a child-sized robot KASPAR, which could move its arms, head, and display facial expressions, to have unconstrained interaction with children. The behaviors of the robot was controlled either by the investigator or the child. They found that children interacted with KASPAR using touch and gaze behaviors and these social skills appeared to generalize to the co-present investigator. The general approach for a robot to elicit and/or teach social behavior is to involve individuals with ASD in either free-form interaction or structured interaction. In free-form interaction, individuals with ASD are allowed to play with the robot as they wish and detected social behaviors are rewarded through sensory feedback and encouragement. Structured interaction typically has the robot play a simple game with individuals with ASD, in which the robot is the social agent that demonstrates social behaviors and guides or enforces the individuals to interact socially with it. Some researchers have developed SAR systems to target more than one skill, such as Boccanfuso and O’Kane [107] who developed CHARLIE to teach imitation, turn-taking, and self-initiated interaction.

Eventually, the objective is not to improve an individual’s social interaction and communication with the robot, but to generalize the learned skills from robots to real world people. Data from several studies showed that the presence of robots helps elicit social interaction between a child and the therapist or experimenter [42, 100, 108, 109]. This observation motivated the development of SAR to support triadic interaction where two individuals interact with one robot, or an individual interacts with the robot and a therapist. The research on triadic interaction is still in its early stage and there are only a few works in the literature. Billard et al. [108] developed Robota to play an imitation game with two children or with one child and an experimenter. The role of the Robota is more of a mediator or an object of shared attention rather than providing direct guidance to foster social interaction. Costa et al. [104] developed a SAR system to kick a ball with one child and the robot was controlled by another child. Robot CHARLIE [107] involved two children in cooperative imitation play by having the robot acted as a mediator to pass the pose from one child to another. In these designs, each child only interacted with the robot. The difference between this type of triadic interaction and one-to-one interaction is that the child is aware of the presence of another child and how their own actions would affect another child’s interaction with the robot. Finally, Wainer et al. [43] used robot KASPAR to play a triadic imitation game with two children. Children and the robot took turns to direct a gesture and the rest of the players followed the gesture. The robot encouraged social

interaction between the children by gazing towards the children and providing feedback when the children did not follow each other.

### *1.3.2.2 Models of adaptation*

Despite the reported benefits of robot-mediated therapies for individuals with ASD, many of these SAR systems were controlled remotely by an experimenter and some had only open-loop interactive capabilities. A few SAR systems were developed with the ability to automatically register a child's behaviors of interest. These sensing modules can interpret behaviors with a variety of complexity. Simple behavior detection includes hand and face detection and tracking [107, 108], tactile event detection [53], and proximity detection [51]. Complex behavior detection includes arm pose detection [50], single gesture detection [96], mixed gesture detection [97], and large range gaze estimation and tracking [52]. These sensing modules allow SAR to understand the explicit behaviors of individuals with ASD. Fewer studies investigated models to interpret implicit mental states hidden behind behavioral performance. The robot FACE used physiological signals and eye gaze to predict the emotional reactions from individuals with ASD [105]. Feil-Seifer et al. estimated the interaction between a robot and a child as either positive or negative based on their distance history [51]. Liu et al. [54] adapted robot behavior based on individual preference as captured using physiological signals. Francois et al. [110] recognized a child's tactile interaction styles in terms of gentleness and frequency based on tactile interaction history.

Similar to SAR for older adults, models of interaction are generally task-specific. Levels of robot prompts were handcrafted and the majority followed "least-to-most" hierarchical protocol as higher levels of stimulation were found to be better than lower levels of stimulation [95]. Only one work attempted to design generalized models of interaction. Feil-Seifer and Mataric proposed B<sup>3</sup>IA (Behavior-Based Behavior Intervention Architecture) for robot-mediated ASD intervention [111]. B<sup>3</sup>IA is a behavior-based control architecture for the control of SAR that is applicable for reuse across a variety of intervention tasks and scenarios.

Models of robots are the least investigated aspect in this field. Greczek et al. [50] has attempted to develop a computational model that could use the child's response to learn his/her task ability level so that the robot could adjust its prompt to match with the child's ability level. However, this model was not executed properly in their SAR system due to the fact that the model could not get enough data to adjust robot prompt levels. This indicates that the model could not learn based on very small sample size quickly.

### **1.3.3 Intelligent HCI systems for people with special needs**

In this section, I highlight the HCI systems that are relevant for my research and those that motivated the work on models of machines. A considerable amount of literature has been published on affective and cognitive states recognition. Liu et al. [112] designed two computer-based cognitive tasks in order to elicit three affective states that are important in ASD intervention, which were anxiety, engagement, and liking. A large set of physiological data were collected as children were involved in the tasks, including electrocardiogram, impedance cardiogram, photoplethysmogram, heart sound, skin conductance, electromyogram, and peripheral temperature. These physiological data were then linked to the affective states using a support vector machine-based affective model. Other researchers used physiological data for detecting affective states in other tasks such as in a virtual reality-based social communication task to estimate anxiety level [113], and in a virtual reality-based driving task to estimate engagement, enjoyment, frustration, and boredom [114]. Eye gaze is another physiological data applied for affective and cognitive states recognition. Lahiri et al. [115] used eye gaze metrics such as gaze pattern, pupil dilation, and blink rate to predict subject's engagement in a virtual reality-based social communication task. Zhang et al. [55] measured the cognitive states of adolescents with ASD in a driving task based on their eye gaze data. Zhang et al. [116] also combined peripheral physiological signals, eye gaze, and driving task performance for the purpose of task difficulty level adjustment. Since individuals with ASD have atypical eye gaze patterns that have an effect on meaningful skills such as facial expression recognition [117] and driving [118], a growing number of researchers developed HCI systems that adapted to the eye gaze patterns of these individuals [119, 120].

In terms of models of machines, Hoey et al. [121] developed the COACH prompting system to assist older adults with dementia to perform handwashing task. The COACH prompting system had five levels of prompting actions, which were: do nothing, minimal prompts, maximal prompts, video prompts, and call for human assistance. The handwashing task was modelled as a partially observable Markov decision process (POMDP) that has task-related states and three attitude states (dementia level, awareness, and responsiveness). The attitude states allowed the system to learn its interaction with older adults in addition to their task performance such as whether the older adult is aware of the task and whether the older adults would respond to a certain prompt. With information from task performance and attitude, the COACH system could automatically select appropriate prompts to guide older adults through handwashing. The COACH system was tested with six older adults with moderate-to-severe dementia. Each older adults completed 40 trials in 8 weeks. These older adults had 11% more independent handwashing steps and 60% fewer interactions with a human caregiver with the help of COACH [17]. Recently, the COACH system was extended by Lin et al. [122] and Robillard and Hoey [123] to incorporate emotional response from

older adults measured by their body posture. Different emotional responses would trigger personalized behavior and emotion in the COACH system following the affect control theory in the sociological literature, therefore making the system affect-sensitive and act more socially.

#### **1.3.4 Summary**

The literature review identified the needs to further investigate mental state models of people, to develop SAR systems for many-to-one interaction, to design more generalized models of interaction, and to design models of machines. In particular:

*Models of people* Many of the previous works used physiological data including peripheral physiological signals and eye gaze data to learn affective and cognitive states of people with special needs. A few researchers explored other sensory channels such as body posture, tactile event, and verbal intonation. Electroencephalogram (EEG) measures the electrical activity over the scalp that carries rich information of human brain. In HMI, EEG devices are bulky and expensive and are mainly used as a new control and communication channel for individuals with severe motor disabilities [124]. With the emergence of more affordable and lightweight commercial EEG devices such as Emotiv EPOC (founded in 2011), it becomes possible to learn models of people from EEG signals in order to build a brain machine interface to enrich HMI. This type of data-driven model based on EEG signals have not been investigated.

*Models of interaction* Most SAR systems to date have predominantly focused on one-to-one interaction. Many-to-one interaction has the benefits of alleviating social isolation or loneliness, and fostering social interaction and communication. Thus it is important to develop multi-user SAR systems to support both HRI and HHI. Almost all the previous work were designed within the scope of a particular interaction scenario to engage older adults in a physical or cognitive activity or to teach individuals with ASD a single skill. As a consequence, the models of interaction are task-specific and cannot be reused for a different interaction scenario. This calls for the design of a more generalized model of interaction.

*Models of machines* An adaptive system not only needs to respond autonomously to human behaviors, but also needs to adapt based on human responses to robot behaviors. However, there is very limited work on models of machines that enable the robot to understand how its behaviors affect human behaviors. In this context, human behaviors are recognized by models of people. However, implicit mental state models of people is less studied in human intention. Human beings are very good at reasoning about other's intention. Intention is context-sensitive and could interpret both mental state and task-related information. Perhaps the most useful intention recognition is through eye gaze to infer regions of interest. Therefore, it

is important to design SAR systems to learn human intention and to learn how its behavior impact human intention and adapt accordingly.

## **1.4 Structure of this Dissertation**

My research focuses on the mental state models of people, the design and development of SAR systems for both one-to-one and many-to-one HRI with older adults, and the design of model of interaction and model of machine. The models of people, interaction, and machine are building blocks for adaptive automation of SAR systems. In this research, the SAR systems for older adults were developed to deliver multimodal therapies with an emphasis on social engagement. For ASD intervention, mental state models were built in order to enable personalized HMI.

In CHAPTER 2, I describe my work on developing data-driven models of people to capture five mental states that are relevant to driving skill training for individuals with ASD. The five mental states that I investigated and trained models on were engagement, enjoyment, boredom, frustration, and mental workload. These models were built using participants' passive EEG signals recorded while they were performing the driving task. Results implied that models based on EEG activations can detect with high accuracy the states of low engagement, low enjoyment, high frustration, and high workload for ASD population. Boredom recognition had relatively low accuracy.

In CHAPTER 3, I present my work on the design of a generalized model of interaction based on engagement models of one to many users, and the developments and user studies of two SAR systems, a one-to-one interaction with five activities and a triadic interaction with a single activity. EEG signals were analyzed offline to estimate older adults' engagement intention variable in the engagement models. Results indicated that both SAR systems were positively accepted by older adults with and without cognitive impairment, the generalized model can be used for one-to-one and many-to-one HRI, and the selection of the EEG feature has the potential for objectively measuring older adults' engagement intention.

In CHAPTER 4, I describe my work on the design and development of an autonomous robot-mediated interaction system to foster social interaction among older adults within a multimodal task. This system consisted of three major components, which were i) a multimodal task with embedded physical, cognitive, and social stimuli; ii) a robot control mechanism to keep older adults engaged in both HRI and HHI; and iii) data analysis algorithms to quantify older adults' social interaction and activity engagement. Results indicated that this system could involve two older adults to perform multimodal activities, could engage them in HRI and HHI, and could quantitatively measure their social interaction and activity engagement. Changes of participants' activity engagement and social interaction within a single session were positive

and thus indicated the potential usefulness of the system and supported further investigation of the efficacy of such systems by conducting multi-session experiments.

In CHAPTER 5, I present my work on conducting a multi-session triadic HRI experiment in real world setting with older adults residing at local retirement communities. The SAR systems for triadic interaction developed in CHAPTER 3 and CHAPTER 4 were combined to create Ro-Tri and tested with seven pairs of older adults. Both subjective and objective data were collected to gather feasibility data. Results indicated that older adults' visual attention towards their peers during HRI improved slightly from session one to session six, their interest, perception, and engagement in the robot-mediated activities were either maintained or slightly improved. Results also demonstrate the ability of gathered data to assess changes of older adults' engagement and physiological indicators.

In CHAPTER 6, I describe my work on the design of a novel mathematical model of adaptation for multi-user HRI. This work formally modelled multi-user HRI with an integrated model combining a model of people, a model of interaction, and a model of machine that takes into account individual differences and is applicable for reuse across a variety of interaction scenarios. Simulation was conducted with a concrete multi-user HRI scenario that was modified from the *RockSample* problem. Strategic behaviors of humans in the *RockSample* game was designed to condition on user's level of cooperation with other users and with the robot, and user's noise level. The simulator was implemented using reinforcement learning and did not have information about the formalized model for multi-user HRI. Simulation results demonstrated the ability of the model to estimate and shape HA, HI, and HCL, learn human adaptability, and facilitate task completion.

Finally, CHAPTER 7 summarizes the primary contributions of the dissertation research, in terms of technical contributions and contributions to the science of elder care and ASD intervention, and highlights future directions.

## REFERENCES

- [1] G. Johannsen, "Human-machine interaction," *Control Systems, Robotics and Automation*, vol. 21, pp. 132-62, 2009.
- [2] M. A. Goodrich and A. C. Schultz, "Human-robot interaction: a survey," *Foundations and trends in human-computer interaction*, vol. 1, pp. 203-275, 2007.
- [3] A. Jaimes and N. Sebe, "Multimodal human-computer interaction: A survey," *Computer vision and image understanding*, vol. 108, pp. 116-134, 2007.
- [4] Federal Interagency Forum on Aging-Related Statistics, "Older Americans 2016: Key Indicators of Well-Being," *Federal Interagency Forum on Aging-Related Statistics*, 2016.
- [5] US Department of Health and Human Services, "Multiple chronic conditions—a strategic framework: optimum health and quality of life for individuals with multiple chronic conditions," *Washington, DC: US Department of Health and Human Services*, vol. 2, 2010.
- [6] A. Fiske, J. L. Wetherell, and M. Gatz, "Depression in older adults," *Annual review of clinical psychology*, vol. 5, pp. 363-389, 2009.
- [7] S. M. Friedman and D. A. Mendelson, "Epidemiology of fragility fractures," *Clinics in geriatric medicine*, vol. 30, pp. 175-181, 2014.
- [8] Alzheimer's Association, "2018 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 14, pp. 367-429, 2018.
- [9] Centers for Disease Control and Prevention, "Chronic diseases: The leading causes of death and disability in the United States," *Chronic Disease Overview*, 2015.
- [10] Alzheimer's Association, "2017 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 13, pp. 325-373, 2017.
- [11] B. C. Spillman, "Why do elders receiving informal home care transition to long stay nursing home residency," *Report to Department of Health and Human Services, Assistant Secretary for Planning and Evaluation, Office of Disability, Aging and Long-Term Care Policy, Contract HHSP23337033T*, 2014.
- [12] G. Livingston, A. Sommerlad, V. Orgeta, S. G. Costafreda, J. Huntley, D. Ames, C. Ballard, S. Banerjee, A. Burns, and J. Cohen-Mansfield, "Dementia prevention, intervention, and care," *The Lancet*, vol. 390, pp. 2673-2734, 2017.
- [13] A. L. Bossen, H. Kim, K. N. Williams, A. E. Steinhoff, and M. Strieker, "Emerging roles for telemedicine and smart technologies in dementia care," *Smart homecare technology and telehealth*, vol. 3, p. 49, 2015.
- [14] T. Suzuki, S. Murase, T. Tanaka, and T. Okazawa, "New approach for the early detection of dementia by recording in-house activities," *Telemedicine and e-Health*, vol. 13, pp. 41-44, 2007.
- [15] T. L. Hayes, F. Abendroth, A. Adami, M. Pavel, T. A. Zitzelberger, and J. A. Kaye, "Unobtrusive assessment of activity patterns associated with mild cognitive impairment," *Alzheimer's & dementia: the journal of the Alzheimer's Association*, vol. 4, pp. 395-405, 2008.
- [16] P. Rashidi and A. Mihailidis, "A survey on ambient-assisted living tools for older adults," *IEEE journal of biomedical and health informatics*, vol. 17, pp. 579-590, 2013.
- [17] A. Mihailidis, J. N. Boger, T. Craig, and J. Hoey, "The COACH prompting system to assist older adults with dementia through handwashing: An efficacy study," *BMC geriatrics*, vol. 8, p. 28, 2008.
- [18] T. Shibata and K. Wada, "Robot therapy: a new approach for mental healthcare of the elderly—a mini-review," *Gerontology*, vol. 57, pp. 378-386, 2011.
- [19] S. Coradeschi, A. Cesta, G. Cortellessa, L. Coraci, J. Gonzalez, L. Karlsson, F. Furfari, A. Loutfi, A. Orlandini, and F. Palumbo, "Giraffplus: Combining social interaction and long term monitoring for promoting independent living," in *Human system interaction (HSI), 2013 the 6th international conference on*, 2013, pp. 578-585.
- [20] D. McColl, W. Y. G. Louie, and G. Nejat, "Brian 2.1: A socially assistive robot for the elderly and cognitively impaired," *Robotics & Automation Magazine, IEEE*, vol. 20, pp. 74-83, 2013.



- [21] N. Alm, R. Dye, G. Gowans, J. Campbell, A. Astell, and M. Ellis, "A communication support system for older people with dementia," *Computer*, vol. 40, 2007.
- [22] M. Ienca, J. Fabrice, B. Elger, M. Caon, A. S. Pappagallo, R. W. Kressig, and T. Wangmo, "Intelligent assistive technology for Alzheimer's disease and other dementias: a systematic review," *Journal of Alzheimer's Disease*, vol. 56, pp. 1301-1340, 2017.
- [23] M.-V. Zunzunegui, B. E. Alvarado, T. Del Ser, and A. Otero, "Social networks, social integration, and social engagement determine cognitive decline in community-dwelling Spanish older adults," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 58, pp. S93-S100, 2003.
- [24] W. J. Chodzko-Zajko, D. N. Proctor, M. A. F. Singh, C. T. Minson, C. R. Nigg, G. J. Salem, and J. S. Skinner, "Exercise and physical activity for older adults," *Medicine & science in sports & exercise*, vol. 41, pp. 1510-1530, 2009.
- [25] M. Cattan, M. White, J. Bond, and A. Learmouth, "Preventing social isolation and loneliness among older people: a systematic review of health promotion interventions," *Ageing & Society*, vol. 25, pp. 41-67, 2005.
- [26] C. Groot, A. Hooghiemstra, P. Raijmakers, B. Van Berckel, P. Scheltens, E. Scherder, W. Van der Flier, and R. Ossenkoppele, "The effect of physical activity on cognitive function in patients with dementia: a meta-analysis of randomized control trials," *Ageing research reviews*, vol. 25, pp. 13-23, 2016.
- [27] E. Aguirre, R. T. Woods, A. Spector, and M. Orrell, "Cognitive stimulation for dementia: a systematic review of the evidence of effectiveness from randomised controlled trials," *Ageing research reviews*, vol. 12, pp. 253-262, 2013.
- [28] N. Farina, J. Rusted, and N. Tabet, "The effect of exercise interventions on cognitive outcome in Alzheimer's disease: a systematic review," *International Psychogeriatrics*, vol. 26, pp. 9-18, 2014.
- [29] J. Cohen-Mansfield, M. S. Marx, M. Dakheel-Ali, and K. Thein, "The use and utility of specific nonpharmacological interventions for behavioral symptoms in dementia: an exploratory study," *The American Journal of Geriatric Psychiatry*, vol. 23, pp. 160-170, 2015.
- [30] C. Anderson-Hanley, P. J. Arciero, A. M. Brickman, J. P. Nimon, N. Okuma, S. C. Westen, M. E. Merz, B. D. Pence, J. A. Woods, and A. F. Kramer, "Exergaming and older adult cognition: a cluster randomized clinical trial," *American journal of preventive medicine*, vol. 42, pp. 109-119, 2012.
- [31] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders (DSM-5®)*: American Psychiatric Pub, 2013.
- [32] J. Baio, L. Wiggins, D. L. Christensen, M. J. Maenner, J. Daniels, Z. Warren, M. Kurzius-Spencer, W. Zahorodny, C. R. Rosenberg, and T. White, "Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014," *MMWR Surveillance Summaries*, vol. 67, p. 1, 2018.
- [33] Z. Warren, M. L. McPheeters, N. Sathe, J. H. Foss-Feig, A. Glasser, and J. Veenstra-VanderWeele, "A systematic review of early intensive intervention for autism spectrum disorders," *Pediatrics*, vol. 127, pp. e1303-e1311, 2011.
- [34] M. A. Maglione, D. Gans, L. Das, J. Timbie, and C. Kasari, "Nonmedical interventions for children with ASD: Recommended guidelines and further research needs," *Pediatrics*, vol. 130, pp. S169-S178, 2012.
- [35] D. R. Hendricks and P. Wehman, "Transition from school to adulthood for youth with autism spectrum disorders: Review and recommendations," *Focus on Autism and Other Developmental Disabilities*, vol. 24, pp. 77-88, 2009.
- [36] D. Amendah, S. Grosse, G. Peacock, and D. Mandell, "The economic costs of autism: A review," *Autism spectrum disorders*, pp. 1347-1360, 2011.
- [37] D. J. Ricks and M. B. Colton, "Trends and considerations in robot-assisted autism therapy," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, 2010, pp. 4354-4359.

- [38] S. L. Odom, J. L. Thompson, S. Hedges, B. A. Boyd, J. R. Dykstra, M. A. Duda, K. L. Szidon, L. E. Smith, and A. Bord, "Technology-aided interventions and instruction for adolescents with autism spectrum disorder," *Journal of autism and developmental disorders*, vol. 45, pp. 3805-3819, 2015.
- [39] O. Grynszpan, P. L. Weiss, F. Perez-Diaz, and E. Gal, "Innovative technology-based interventions for autism spectrum disorders: a meta-analysis," *Autism*, vol. 18, pp. 346-361, 2014.
- [40] J. J. Diehl, L. M. Schmitt, M. Villano, and C. R. Crowell, "The clinical use of robots for individuals with autism spectrum disorders: A critical review," *Research in autism spectrum disorders*, vol. 6, pp. 249-262, 2012.
- [41] E. M. Whyte, J. M. Smyth, and K. S. Scherf, "Designing serious game interventions for individuals with autism," *Journal of autism and developmental disorders*, vol. 45, pp. 3820-3831, 2015.
- [42] B. Robins, K. Dautenhahn, and P. Dickerson, "From isolation to communication: a case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot," in *Advances in Computer-Human Interactions, 2009. ACHI'09. Second International Conferences on*, 2009, pp. 205-211.
- [43] J. Wainer, B. Robins, F. Amirabdollahian, and K. Dautenhahn, "Using the humanoid robot KASPAR to autonomously play triadic games and facilitate collaborative play among children with autism," *IEEE Transactions on Autonomous Mental Development*, vol. 6, pp. 183-199, 2014.
- [44] L. Millen, S. Cobb, H. Patel, and T. Glover, "Collaborative virtual environment for conducting design sessions with students with autism spectrum conditions," in *Proc. 9th Intl Conf. on Disability, Virtual Reality and Assoc. Technologies*, 2012, pp. 269-278.
- [45] Y. Cheng, H.-C. Chiang, J. Ye, and L.-h. Cheng, "Enhancing empathy instruction using a collaborative virtual learning environment for children with autistic spectrum conditions," *Computers & Education*, vol. 55, pp. 1449-1458, 2010.
- [46] A. Ben-Sasson, L. Lamash, and E. Gal, "To enforce or not to enforce? The use of collaborative interfaces to promote social skills in children with high functioning autism spectrum disorder," *Autism*, vol. 17, pp. 608-622, 2013.
- [47] E. Gal, N. Bauminger, D. Goren-Bar, F. Pianesi, O. Stock, M. Zancanaro, and P. L. T. Weiss, "Enhancing social communication of children with high-functioning autism through a co-located interface," *Ai & Society*, vol. 24, p. 75, 2009.
- [48] J. Fasola and M. J. Matarić, "A Socially Assistive Robot Exercise Coach for the Elderly," *Journal of Human-Robot Interaction*, vol. 2, pp. 3-32, 2013.
- [49] Z. Zheng, E. M. Young, A. R. Swanson, A. S. Weitlauf, Z. E. Warren, and N. Sarkar, "Robot-mediated imitation skill training for children with autism," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, pp. 682-691, 2016.
- [50] J. Greczek, E. Kaszubski, A. Atrash, and M. Matarić, "Graded cueing feedback in robot-mediated imitation practice for children with autism spectrum disorders," in *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*, 2014, pp. 561-566.
- [51] D. Feil-Seifer and M. J. Matarić, "Automated detection and classification of positive vs. negative robot interactions with children with autism using distance-based features," in *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on*, 2011, pp. 323-330.
- [52] Z. Zheng, G. Nie, A. Swanson, A. Weitlauf, Z. Warren, and N. Sarkar, "Longitudinal Impact of Autonomous Robot-Mediated Joint Attention Intervention for Young Children with ASD," in *International Conference on Social Robotics*, 2016, pp. 581-590.
- [53] F. Amirabdollahian, B. Robins, K. Dautenhahn, and Z. Ji, "Investigating tactile event recognition in child-robot interaction for use in autism therapy," in *Engineering in medicine and biology society, EMBC, 2011 annual international conference of the IEEE*, 2011, pp. 5347-5351.
- [54] C. Liu, K. Conn, N. Sarkar, and W. Stone, "Online affect detection and robot behavior adaptation for intervention of children with autism," *IEEE transactions on robotics*, vol. 24, pp. 883-896, 2008.
- [55] L. Zhang, J. Wade, A. Swanson, A. Weitlauf, Z. Warren, and N. Sarkar, "Cognitive state measurement from eye gaze analysis in an intelligent virtual reality driving system for autism

- intervention," in *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*, 2015, pp. 532-538.
- [56] D. Feil-Seifer and M. J. Mataric, "Defining socially assistive robotics," in *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on*, 2005, pp. 465-468.
- [57] L. Tiberio, A. Cesta, G. Cortellessa, L. Padua, and A. R. Pellegrino, "Assessing affective response of older users to a telepresence robot using a combination of psychophysiological measures," in *RO-MAN, 2012 IEEE*, 2012, pp. 833-838.
- [58] H.-M. Gross, C. Schroeter, S. Mueller, M. Volkhardt, E. Einhorn, A. Bley, T. Langner, M. Merten, C. Huijnen, and H. van den Heuvel, "Further progress towards a home robot companion for people with mild cognitive impairment," in *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*, 2012, pp. 637-644.
- [59] K. Wada, T. Shibata, T. Saito, K. Sakamoto, and K. Tanie, "Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, 2005, pp. 2785-2790.
- [60] S. C. Kramer, E. Friedmann, and P. L. Bernstein, "Comparison of the effect of human interaction, animal-assisted therapy, and AIBO-assisted therapy on long-term care residents with dementia," *Anthrozoös*, vol. 22, pp. 43-57, 2009.
- [61] A. V. Libin and E. V. Libin, "Person-robot interactions from the robopsychologists' point of view: The robotic psychology and robototherapy approach," *Proceedings of the IEEE*, vol. 92, pp. 1789-1803, 2004.
- [62] A. Libin and J. Cohen-Mansfield, "Therapeutic robot for nursing home residents with dementia: preliminary inquiry," *American Journal of Alzheimer's Disease & Other Dementias®*, vol. 19, pp. 111-116, 2004.
- [63] R. Bemelmans, G. J. Gelderblom, P. Jonker, and L. De Witte, "Socially assistive robots in elderly care: A systematic review into effects and effectiveness," *Journal of the American Medical Directors Association*, vol. 13, pp. 114-120. e1, 2012.
- [64] 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*, vol. 74, pp. 14-20, 2013.
- [65] W. Moyle, M. Cooke, E. Beattie, C. Jones, B. Klein, G. Cook, and C. Gray, "Exploring the effect of companion robots on emotional expression in older adults with dementia: A pilot randomized controlled trial," *Journal of gerontological nursing*, 2013.
- [66] E. Broadbent, R. Stafford, and B. MacDonald, "Acceptance of healthcare robots for the older population: Review and future directions," *International Journal of Social Robotics*, vol. 1, p. 319, 2009.
- [67] T. Klamer and S. B. Allouch, "Acceptance and use of a social robot by elderly users in a domestic environment," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2010 4th International Conference on-NO PERMISSIONS*, 2010, pp. 1-8.
- [68] R. Looije, M. A. Neerinx, and F. Cnossen, "Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors," *International Journal of Human-Computer Studies*, vol. 68, pp. 386-397, 2010.
- [69] M. E. Pollack, L. Brown, D. Colbry, C. Orosz, B. Peintner, S. Ramakrishnan, S. Engberg, J. T. Matthews, J. Dunbar-Jacob, and C. E. McCarthy, "Pearl: A mobile robotic assistant for the elderly," in *AAAI workshop on automation as eldercare*, 2002, pp. 85-91.
- [70] B. Graf, M. Hans, and R. D. Schraft, "Care-O-bot II—Development of a next generation robotic home assistant," *Autonomous robots*, vol. 16, pp. 193-205, 2004.
- [71] R. Khosla and M.-T. Chu, "Embodying care in Matilda: an affective communication robot for emotional wellbeing of older people in Australian residential care facilities," *ACM Transactions on Management Information Systems (TMIS)*, vol. 4, p. 18, 2013.

- [72] R. Khosla, K. Nguyen, and M.-T. Chu, "Assistive robot enabled service architecture to support home-based dementia care," in *Service-Oriented Computing and Applications (SOCA), 2014 IEEE 7th International Conference on*, 2014, pp. 73-80.
- [73] T. Inoue, M. Nihei, T. Narita, M. Onoda, R. Ishiwata, I. Mamiya, M. Shino, H. Kojima, S. Ohnaka, and Y. Fujita, "Field-based development of an information support robot for persons with dementia," *Technology and Disability*, vol. 24, pp. 263-271, 2012.
- [74] I. Bäck, K. Makela, and J. Kallio, "Robot-Guided Exercise Program for the Rehabilitation of Older Nursing Home Residents," *Annals of Long-Term Care: Clinical Care and Aging*, vol. 21, pp. 38-41, 2013.
- [75] B. Görer, A. A. Salah, and H. L. Akın, "A robotic fitness coach for the elderly," in *International Joint Conference on Ambient Intelligence*, 2013, pp. 124-139.
- [76] K. Yoshino, M. Kouda, and Z. Shanjun, "Correct Motion Advice on Rehabilitation Instruction Robot by Superimposing Instructor CG Model," in *Intelligent Networks and Intelligent Systems (ICINIS), 2012 Fifth International Conference on*, 2012, pp. 333-336.
- [77] P. Gadde, H. Kharrazi, H. Patel, and K. F. MacDorman, "Toward monitoring and increasing exercise adherence in older adults by robotic intervention: a proof of concept study," *Journal of Robotics*, 2011.
- [78] T. Shibano, Y. Ho, Y. Kono, Y. Fujimoto, and T. Yamaguchi, "Daily support system for care prevention by using interaction monitoring robot," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, 2010, pp. 3477-3482.
- [79] Y. Matsusaka, H. Fujii, T. Okano, and I. Hara, "Health exercise demonstration robot TAIZO and effects of using voice command in robot-human collaborative demonstration," in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, 2009, pp. 472-477.
- [80] A. Tapus, "Improving the quality of life of people with dementia through the use of socially assistive robots," in *Advanced Technologies for Enhanced Quality of Life, 2009. AT-EQUAL'09.*, 2009, pp. 81-86.
- [81] M. Kanoh, Y. Oida, Y. Nomura, A. Araki, Y. Konagaya, K. Ihara, T. Shimizu, and K. Kimura, "Examination of practicability of communication robot-assisted activity program for elderly people," *Journal of Robotics and Mechatronics*, vol. 23, p. 3, 2011.
- [82] W. Y. G. Louie, T. Vaquero, G. Nejat, and J. C. Beck, "An autonomous assistive robot for planning, scheduling and facilitating multi-user activities," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, 2014, pp. 5292-5298.
- [83] D. Szafir and B. Mutlu, "Pay attention!: designing adaptive agents that monitor and improve user engagement," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2012, pp. 11-20.
- [84] P. Bhuvaneshwari, S. Vignesh, S. Papitha, and R. Dharmarajan, "Humanoid robot based physiotherapeutic assistive trainer for elderly health care," in *Recent Trends in Information Technology (ICRTIT), 2013 International Conference on*, 2013, pp. 163-168.
- [85] J. Fan, L. Beuscher, P. A. Newhouse, L. C. Mion, and N. Sarkar, "A robotic coach architecture for multi-user human-robot interaction (RAMU) with the elderly and cognitively impaired," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, 2016, pp. 445-450.
- [86] M. Zecca, Y. Mizoguchi, K. Endo, F. Iida, Y. Kawabata, N. Endo, K. Itoh, and A. Takanishi, "Whole body emotion expressions for KOBIAN humanoid robot—preliminary experiments with different Emotional patterns—," in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, 2009, pp. 381-386.
- [87] Y.-H. Wu, J. Wrobel, V. Cristancho-Lacroix, L. Kamali, M. Chetouani, D. Duhaut, B. Le Pénédic, C. Jost, V. Dupourque, and M. Ghriissi, "Designing an assistive robot for older adults: the ROBADMOM project," *IRBM*, vol. 34, pp. 119-123, 2013.

- [88] R. C. Arkin, M. Fujita, T. Takagi, and R. Hasegawa, "An ethological and emotional basis for human–robot interaction," *Robotics and Autonomous Systems*, vol. 42, pp. 191-201, 2003.
- [89] Y. Matsuyama, H. Taniyama, S. Fujie, and T. Kobayashi, "Framework of Communication Activation Robot Participating in Multiparty Conversation," in *AAAI Fall Symposium: Dialog with Robots*, 2010.
- [90] M. Simonov and G. Delconte, "Humanoid assessing rehabilitative exercises," *Methods of information in medicine*, vol. 54, pp. 114-121, 2015.
- [91] H. Gross, C. Schroeter, S. Mueller, M. Volkhardt, E. Einhorn, A. Bley, C. Martin, T. Langner, and M. Merten, "Progress in developing a socially assistive mobile home robot companion for the elderly with mild cognitive impairment," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, 2011, pp. 2430-2437.
- [92] J. Chan and G. Nejat, "Promoting engagement in cognitively stimulating activities using an intelligent socially assistive robot," in *Advanced Intelligent Mechatronics (AIM), 2010 IEEE/ASME International Conference on*, 2010, pp. 533-538.
- [93] J. Chan and G. Nejat, "Social Intelligence for a Robot Engaging People in Cognitive Training Activities," *International Journal of Advanced Robotic Systems*, vol. 9, Oct 2012.
- [94] R. Emanuel and S. Weir, "Catalysing communication in an autistic child in a LOGO-like learning environment," in *Proceedings of the 2nd Summer Conference on Artificial Intelligence and Simulation of Behaviour*, 1976, pp. 118-129.
- [95] P. Pennisi, A. Tonacci, G. Tartarisco, L. Billeci, L. Ruta, S. Gangemi, and G. Pioggia, "Autism and social robotics: A systematic review," *Autism Research*, vol. 9, pp. 165-183, 2016.
- [96] Z. Zheng, S. Das, E. M. Young, A. Swanson, Z. Warren, and N. Sarkar, "Autonomous robot-mediated imitation learning for children with autism," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, 2014, pp. 2707-2712.
- [97] Z. Zheng, E. M. Young, A. Swanson, A. Weitlauf, Z. Warren, and N. Sarkar, "Robot-mediated mixed gesture imitation skill training for young children with ASD," in *Advanced Robotics (ICAR), 2015 International Conference on*, 2015, pp. 72-77.
- [98] A. Duquette, F. Michaud, and H. Mercier, "Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism," *Autonomous robots*, vol. 24, pp. 147-157, 2008.
- [99] S. M. Anzalone, E. Tilmont, S. Boucenna, J. Xavier, A.-L. Jouen, N. Bodeau, K. Maharatna, M. Chetouani, D. Cohen, and M. S. Group, "How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3D+ time) environment during a joint attention induction task with a robot," *Research in autism spectrum disorders*, vol. 8, pp. 814-826, 2014.
- [100] H. Kozima, C. Nakagawa, and Y. Yasuda, "Children–robot interaction: a pilot study in autism therapy," *Progress in Brain Research*, vol. 164, pp. 385-400, 2007.
- [101] T. Esubalew, U. Lahiri, A. R. Swanson, J. A. Crittendon, Z. E. Warren, and N. Sarkar, "A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, pp. 289-299, 2013.
- [102] Z. Zheng, L. Zhang, E. Bekele, A. Swanson, J. A. Crittendon, Z. Warren, and N. Sarkar, "Impact of robot-mediated interaction system on joint attention skills for children with autism," in *Rehabilitation Robotics (ICORR), 2013 IEEE International Conference on*, 2013, pp. 1-8.
- [103] S. Shamsuddin, H. Yussof, L. I. Ismail, S. Mohamed, F. A. Hanapiah, and N. I. Zahari, "Humanoid robot NAO interacting with autistic children of moderately impaired intelligence to augment communication skills," *Procedia Engineering*, vol. 41, pp. 1533-1538, 2012.
- [104] S. Costa, C. Santos, F. Soares, M. Ferreira, and F. Moreira, "Promoting interaction amongst autistic adolescents using robots," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, 2010, pp. 3856-3859.
- [105] D. Mazzei, L. Billeci, A. Armato, N. Lazzeri, A. Cisternino, G. Pioggia, R. Iglizzo, F. Muratori, A. Ahluwalia, and D. De Rossi, "The face of autism," in *RO-MAN, 2010 IEEE*, 2010, pp. 791-796.

- [106] O. Damm, K. Malchus, P. Jaecks, S. Krach, F. Paulus, M. Naber, A. Jansen, I. Kamp-Becker, W. Einhaeuser-Treyer, and P. Stenneken, "Different gaze behavior in human-robot interaction in Asperger's syndrome: An eye-tracking study," in *RO-MAN, 2013 IEEE*, 2013, pp. 368-369.
- [107] L. Boccanfuso and J. M. O'Kane, "CHARLIE: An adaptive robot design with hand and face tracking for use in autism therapy," *International Journal of Social Robotics*, vol. 3, pp. 337-347, 2011.
- [108] A. Billard, B. Robins, J. Nadel, and K. Dautenhahn, "Building Robota, a mini-humanoid robot for the rehabilitation of children with autism," *Assistive Technology*, vol. 19, pp. 37-49, 2007.
- [109] J.-J. Cabibihan, H. Javed, M. Ang, and S. M. Aljunied, "Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism," *International Journal of Social Robotics*, vol. 5, pp. 593-618, 2013.
- [110] D. François, K. Dautenhahn, and D. Polani, "Using real-time recognition of human-robot interaction styles for creating adaptive robot behaviour in robot-assisted play," in *Artificial Life, 2009. ALife'09. IEEE Symposium on*, 2009, pp. 45-52.
- [111] D. Feil-Seifer and M. J. Mataric, "B 3 IA: A control architecture for autonomous robot-assisted behavior intervention for children with Autism Spectrum Disorders," in *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, 2008, pp. 328-333.
- [112] C. Liu, K. Conn, N. Sarkar, and W. Stone, "Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder," *International Journal of Human-Computer Studies*, vol. 66, pp. 662-677, 2008.
- [113] S. Kuriakose and U. Lahiri, "Understanding the psycho-physiological implications of interaction with a virtual reality-based system in adolescents with autism: a feasibility study," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, pp. 665-675, 2015.
- [114] D. Bian, J. W. Wade, A. Swanson, Z. Warren, and N. Sarkar, "Physiology-based Affect Recognition During Driving in Virtual Environment for Autism Intervention," in *PhyCS*, 2015, pp. 137-145.
- [115] U. Lahiri, E. Bekele, E. Dohrmann, Z. Warren, and N. Sarkar, "A physiologically informed virtual reality based social communication system for individuals with autism," *Journal of autism and developmental disorders*, vol. 45, pp. 919-931, 2015.
- [116] L. Zhang, J. W. Wade, D. Bian, A. Swanson, Z. Warren, and N. Sarkar, "Data fusion for difficulty adjustment in an adaptive virtual reality game system for autism intervention," in *International Conference on Human-Computer Interaction*, 2014, pp. 648-652.
- [117] M. Uljarevic and A. Hamilton, "Recognition of emotions in autism: a formal meta-analysis," *Journal of autism and developmental disorders*, vol. 43, pp. 1517-1526, 2013.
- [118] B. Reimer, R. Fried, B. Mehler, G. Joshi, A. Bolfek, K. M. Godfrey, N. Zhao, R. Goldin, and J. Biederman, "Brief report: Examining driving behavior in young adults with high functioning autism spectrum disorders: A pilot study using a driving simulation paradigm," *Journal of autism and developmental disorders*, vol. 43, pp. 2211-2217, 2013.
- [119] J. Wade, L. Zhang, D. Bian, J. Fan, A. Swanson, A. Weitlauf, M. Sarkar, Z. Warren, and N. Sarkar, "A gaze-contingent adaptive virtual reality driving environment for intervention in individuals with autism spectrum disorders," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 6, p. 3, 2016.
- [120] E. Bekele, J. Wade, D. Bian, J. Fan, A. Swanson, Z. Warren, and N. Sarkar, "Multimodal adaptive social interaction in virtual environment (MASI-VR) for children with Autism spectrum disorders (ASD)," in *Virtual Reality (VR), 2016 IEEE*, 2016, pp. 121-130.
- [121] J. Hoey, P. Poupart, A. von Bertoldi, T. Craig, C. Boutilier, and A. Mihailidis, "Automated handwashing assistance for persons with dementia using video and a partially observable Markov decision process," *Computer vision and image understanding*, vol. 114, pp. 503-519, 2010.

- [122] L. Lin, S. Czarnuch, A. Malhotra, L. Yu, T. Schröder, and J. Hoey, "Affectively aligned cognitive assistance using Bayesian affect control theory," in *International Workshop on Ambient Assisted Living*, 2014, pp. 279-287.
- [123] J. M. Robillard and J. Hoey, "Emotion and Motivation in Cognitive Assistive Technologies for Dementia," *Computer*, vol. 51, pp. 24-34, 2018.
- [124] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, pp. 1211-1279, 2012.

## CHAPTER 2

### EEG-BASED AFFECT AND WORKLOAD RECOGNITION<sup>1</sup>

#### 2.1 Brief Summary

Many individuals with ASD fail to achieve typical milestones related to adult independence. Driving is one such task that individuals with ASD find it to be particularly challenging. Virtual reality (VR)-based driving skill training provides a safe learning environment that can be designed to optimally engage individuals with ASD. By measuring the affective states and the mental states of individuals with ASD and purposefully adapting the VR-based driving intervention system to keep them in the flow, the VR-based driving intervention may be optimally impactful.

We integrated an electroencephalogram (EEG) sensory modality into a previously designed VR-based driving simulator in order to build data-driven group-level models that could detect the states of low engagement, low enjoyment, high boredom, high frustration, and high workload of individuals with ASD during driving skill training. To demonstrate the feasibility of building such models with high detection accuracy, we developed a two-step feature calibration method to dramatically reduce the training sessions needed compared to individualized model training. We further systematically evaluated feature generation approaches, and evaluated discriminative features in terms of feature and electrode usage.

Leave-one-subject-out nested cross-validation method was applied to evaluate different variations of k-nearest neighbor algorithm, different feature types, and different number of selected discriminative features. The best performing models scored 0.95 over 6 subjects for engagement, 0.895 over 13 subjects for enjoyment, 0.78 over 10 subjects for boredom, 0.875 over 18 subjects for frustration, and 0.855 over 18 subjects for workload. These results are comparable to the extant literature. This work also provided insight on the performance of several different feature types and different electrodes. The power features from bins and HOC-based features performed the best and the most discriminative features were found to be extracted from frontal electrodes. The developed models provide a basis for an EEG-based passive brain computer interface system that has the potential to benefit individuals with ASD with an affect- and workload-based individualized driving skill training intervention.

---

<sup>1</sup> © 2018 IEEE. Reprinted, with permission, from Jing Fan, Joshua Wade, Alexandra Key, Zachary Warren, and Nilanjan Sarkar, EEG-based affect and workload recognition in a virtual driving environment for ASD intervention, Biomedical Engineering, IEEE Transactions on, Jan. 2018.



There were 8 research papers published on this work or related EEG work.

- 1) **Fan, Jing**, Joshua W. Wade, Alexandra P. Key, Zachary E. Warren, and Nilanjan Sarkar. "EEG-based affect and workload recognition in a virtual driving environment for ASD intervention." *IEEE Transactions on Biomedical Engineering* 65, no. 1 (2018): 43-51.
- 2) **Fan, Jing**, Esube Bekele, Zachary Warren, and Nilanjan Sarkar. "EEG analysis of facial affect recognition process of individuals with ASD performance prediction leveraging social context." In *Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), 2017 Seventh International Conference on*, pp. 38-43. IEEE, 2017.
- 3) Zhang, Lian, Joshua Wade, Dayi Bian, **Jing Fan**, Amy Swanson, Amy Weitlauf, Zachary Warren, and Nilanjan Sarkar. "Cognitive load measurement in a virtual reality-based driving system for autism intervention." *IEEE transactions on affective computing* 8, no. 2 (2017): 176-189.
- 4) Wade, Joshua, Lian Zhang, Dayi Bian, **Jing Fan**, Amy Swanson, Amy Weitlauf, Medha Sarkar, Zachary Warren, and Nilanjan Sarkar. "A gaze-contingent adaptive virtual reality driving environment for intervention in individuals with autism spectrum disorders." *ACM Transactions on Interactive Intelligent Systems (TiiS)* 6, no. 1 (2016): 3.
- 5) Bekele, Esubalew, J. Wade, Dayi Bian, **Jing Fan**, Amy Swanson, Zachary Warren, and Nilanjan Sarkar. "Multimodal adaptive social interaction in virtual environment (MASI-VR) for children with Autism spectrum disorders (ASD)." In *Virtual Reality (VR), 2016 IEEE*, pp. 121-130. IEEE, 2016.
- 6) **Fan, Jing**, Joshua W. Wade, Dayi Bian, Alexandra P. Key, Zachary E. Warren, Lorraine C. Mion, and Nilanjan Sarkar. "A Step towards EEG-based brain computer interface for autism intervention." In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp. 3767-3770. IEEE, 2015.
- 7) Wade, Joshua, Dayi Bian, **Jing Fan**, Lian Zhang, Amy Swanson, Medha Sarkar, Amy Weitlauf, Zachary Warren, and Nilanjan Sarkar. "A virtual reality driving environment for training safe gaze patterns: application in individuals with ASD." In *International Conference on Universal Access in Human-Computer Interaction*, pp. 689-697. Springer, Cham, 2015.
- 8) Zhang, Lian, Joshua Wade, Dayi Bian, **Jing Fan**, Amy Swanson, Amy Weitlauf, Zachary Warren, and Nilanjan Sarkar. "Multimodal fusion for cognitive load measurement in an adaptive virtual reality driving task for autism intervention." In *International Conference on Universal Access in Human-Computer Interaction*, pp. 709-720. Springer, Cham, 2015.

## 2.2 Abstract

*Objective:* To build group-level classification models capable of recognizing affective states and mental workload of individuals with autism spectrum disorder (ASD) during driving skill training. *Methods:* Twenty adolescents with ASD participated in a six-session virtual reality driving simulator based experiment, during which their electroencephalogram (EEG) data were recorded alongside driving events and a therapist's rating of their affective states and mental workload. Five feature generation approaches including statistical features, fractal dimension features, higher order crossings (HOC)-based features, power features from frequency bands, and power features from bins ( $\Delta f = 2$  Hz) were applied to extract relevant features. Individual differences were removed with a two-step feature calibration method. Finally, binary classification results based on the k-nearest neighbors algorithm and univariate feature selection method were evaluated by leave-one-subject-out nested cross-validation to compare feature types and identify discriminative features. *Results:* The best classification results were achieved using power features from bins for engagement (0.95) and boredom (0.78), and HOC-based features for enjoyment (0.90), frustration (0.88), and workload (0.86). *Conclusion:* Offline EEG-based group-level classification models are feasible for recognizing binary low and high intensity of affect and workload of individuals with ASD in the context of driving. However, while promising the applicability of the models in an online adaptive driving task requires further development. *Significance:* The developed models provide a basis for an EEG-based passive brain computer interface system that has the potential to benefit individuals with ASD with an affect- and workload-based individualized driving skill training intervention.

## 2.3 Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental syndrome that affects an estimated 1 in 59 children in the US [1] and is the fastest-growing developmental disability. Primary symptoms of ASD include deficits in social interaction, language and communication skills, and restricted, repetitive behaviors [2]. In addition to these core deficit areas, recent evidence suggests that adolescents and young adults with ASD have difficulty in learning safe driving skills [3-5]. In particular, compared with their typically developed peers, individuals with ASD demonstrated unsafe gaze patterns and higher levels of anxiety when operating a driving simulator [6, 7], responded slower during steering, identified fewer social hazards, and showed problematic multi-tasking ability [8, 9]. In the US, driving plays a critical role in everyday life and is essential for achieving adult independence. Given the heterogeneity and developmental nature of ASD [10], effective driving interventions tailored to specific individuals are needed for this population.

While virtual reality (VR)-based intervention for teaching social skills to children with ASD has been investigated in recent years due to various advantages of VR [11, 12], exploration into VR-based driving skill training for adolescents with ASD is only beginning to emerge [7, 13]. VR provides a safe driving skill training environment that can be designed to optimally engage individuals with ASD. Studies of game-based learning environments have argued for the importance of combining game design with flow theory to achieve optimal experience and enhance learning [14, 15]. Flow theory asserts that optimal experience is gained when the challenge level matches the skill level of a player [16]. Hence, by measuring the affective states and the mental workload of individuals with ASD and purposefully adapting the VR-based driving intervention system to keep them in the flow, the VR-based driving intervention may be optimally impactful. As a first step to designing such an individualized intervention system, in this study, an electroencephalogram (EEG) sensory modality was integrated into a VR-based driving simulator to build models for recognizing several affective states and the mental workload of individuals with ASD when they performed driving tasks. The goal of this research is to demonstrate that EEG-based affect and mental workload recognition is feasible during driving in a VR-based simulator so that in the future such an ability can help individualize the training.

In recent years, there has been an increasing interest in developing EEG-based passive brain computer interface (BCI) applications to enrich human-machine interaction. Kohlmorgen et al. trained an individualized mental workload detector using EEG in a real world driving scenario. The workload detector was then applied in real time to switch off the secondary task in the case of high workload [17]. Wang et al. proposed an online closed-loop lapse detection and mitigation system that continuously monitored a driver's EEG signature of fatigue based on EEG spectra, and delivered warnings accordingly during an event-related lane-keeping task using a VR-based driving simulator [18]. Dijksterhuis et al. classified the mental workload of drivers with varying speed- and lane-keeping demand by applying common spatial pattern and a linear discriminant analysis algorithm, again with a driving simulator [19]. Compared with mental workload, affective states are less studied in driving because affective states are not directly related to the safety-critical aspect of driving. Instead, most studies focused on drivers' states such as fatigue, drowsiness, stress, and alertness. Nonetheless, in learning and intelligent systems, EEG-based engagement indices and emotional states recognition have been evaluated and studied by many researchers [20-24]. Various feature generation approaches as well as machine learning algorithms have been applied to improve the reliability of EEG-based affective states recognition [25, 26].

In addition to a paucity of research on EEG-based affective states recognition for driving, the research to date has tended to focus on healthy adults rather than on individuals with greater potential for unsafe driving. Differences in EEG activity between individuals with ASD and their typically developed peers

have been well documented by researchers [27, 28]. Given that driving is a necessary skill for independent living in the US and that individuals with ASD demonstrate a pattern of unsafe driving habits, there is a need to understand how driving skill training can be imparted to these individuals in a safe and flexible environment such as in a VR-based simulator. We believe that such training will be more effective if the system can tailor individual learning experiences based on their affective states and mental workload to accommodate for individual differences inherent in this spectrum disorder.

The primary contributions of this paper are: a) an experimental design to generate EEG data from adolescents with ASD during realistic VR-based driving tasks; b) development of a two-step feature calibration method to allow group-level training. This will dramatically reduce the training sessions needed compared to individualized model training; c) systematic evaluation of feature generation approaches to demonstrate the possibility of group-level affect and workload recognition based on EEG data; and d) systematic evaluation of feature and electrode usage to identify discriminative features. Together they provide a proof of concept that such EEG-based recognition could be useful to individualize ASD intervention. Although existing feature generation methods were applied in the current work, the analyses on EEG data collected from real world tasks with ASD population were not reported in the literature. Such analysis is needed prior to designing an EEG-based BCI for individualized ASD intervention. This paper substantially extends our earlier short conference paper [29] by incorporating rigorous methodology, additional data, and extended results and discussion.

The paper is organized as follows. Section 2.4 describes the VR-based driving simulator and data acquisition modules. Section 2.5 presents the methodology used to systematically analyze the EEG signals. Classification and feature selection results are reported in Section 2.6. In the remaining Sections, we discuss the results and summarize the major findings and significance of the work.

## **2.4 System Description and Data Acquisition**

The VR-based driving system was comprised of a VR driving module and four data acquisition modules, which were used to record EEG, peripheral physiology, eye gaze, and observer rating data (Figure 2-1). The VR driving module consisted of two components: a virtual driving environment rendered from the viewpoint of the driver's perspective and a Logitech G27 driving controller for intuitive control of the virtual vehicle via a steering wheel and a pedal board.

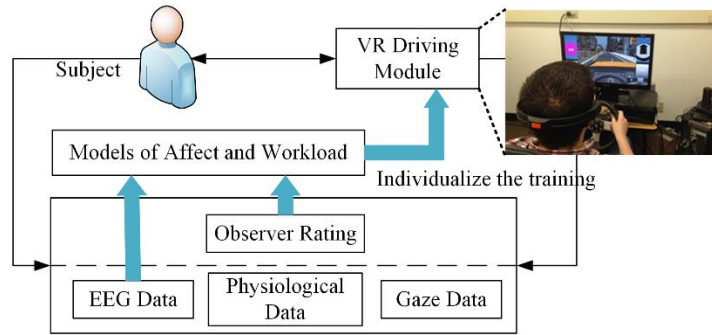


Figure 2-1. Framework Overview

### 2.4.1 Virtual driving environment

The virtual driving environment was created using CityEngine and Autodesk Maya modeling software. The model offered roughly 120 square miles of diverse terrain and provided foundation for enough unique roadways to design hours of driving tasks. Trees, houses, and skyscrapers populated the virtual world, and traffic lights, pedestrians, and various automobiles were used to simulate a bustling city. The roadways included one- and two-way streets as well as an eight lane highway encircling the entire city. Driving tasks, or trials, were designed to utilize each of these particular aspects of the city. Four categories of trials were implemented: turning, merging, speed-maintenance, and laws. Turning trials involved either a left or right turn at an intersection; merging trials included lane changes, overtaking vehicles, and exiting/entering highways; speed-maintenance simply dealt with adjusting speed as appropriate to the specific situation (e.g., highway or school zone speed changes); and laws trials included scenarios in which the driver must obey important road laws, for example, yielding to pedestrians and stopping at stop signs. In all, 144 trials were created.

Eight trials were grouped together in a sequence to create an *assignment*. Subjects attempted to complete an assignment with as few errors as possible. A sufficient number of errors meant the failure and termination of the assignment. Trial errors were monitored by the system for a wide variety of possible offenses. These included – to list only a few – running red lights or stop signs, wrong turns, excessive speed, vehicle collisions, driving in the wrong lane, and failing to move out of the way of emergency vehicles. Errors detected by the system resulted in the subject’s virtual vehicle being reset to the start of the unsuccessful trial. Resets were accompanied by matching audio and text feedback indicating what went wrong and how to avoid the error moving forward. During trials, the system’s built-in navigation system directed drivers – visually and with audio (e.g., “left turn ahead” and “turn left”) – towards their destinations. In addition to the navigation system, the subject’s perspective included a first-person view of the road,

speedometer, turn signal indicators, steering wheel, side view mirrors, and a score shown as text to indicate progress.

In order to modulate the level of challenge faced by the subjects, a range of difficulty levels were introduced. Six different difficulty levels were designed. Several parameters were used to configure the level of difficulty presented by the system: number of vehicles on the road, aggressiveness of other drivers, visibility, weather conditions, and responsiveness of the input device. The easiest level of difficulty was intended to be effortless for most drivers, whereas the hardest level of difficulty was intended to be overly challenging; the other difficulty levels were interpolations between these extremes. These difficulty levels were approved by trained ASD clinicians.

#### **2.4.2 Data acquisition**

We used a 14-channel wireless Emotiv EPOC neuroheadset ([www.emotiv.com](http://www.emotiv.com)) to record EEG signals from locations AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4, defined by the 10-20 system of electrode placement [30]. The reference sensors were placed at locations P3 and P4. The bandwidth of the headset was 0.2-45 Hz and the sampling rate was 128 Hz. We modified an existing EEG data acquisition application to log EEG signals, sensor contact quality, and driving event messages received from the VR driving module. Driving event messages – assignment start/end messages in particular – were used to align EEG signals with driving tasks and observer rating data. Subjects' eye gaze data and physiological data were collected using a Tobii X120 eye tracker ([www.tobii.com](http://www.tobii.com)) and a Biopac MP150 physiological data acquisition system ([www.biopac.com](http://www.biopac.com)), respectively.

We relied on an experienced therapist to report the ground truth of subjects' affective states as well as mental workload during driving. The rating categories employed were engagement, enjoyment, frustration, boredom, and task difficulty. In the context of computer-based learning environments, affective states of engagement, enjoyment, frustration, and boredom have been identified to capture useful learning experience across different learning situations and learners [31]. Therefore, we chose to build a model of affect that was able to recognize these four affective states. The last rating category, task difficulty, was adopted to represent subjects' mental workload. Mental workload characterizes the demands imposed on human's working memory by tasks. It has three attributes, which are mental load, mental effort, and task performance [32, 33]. Task difficulty contributes to mental load. More difficult tasks impose higher demands on the limited mental resources for information processing. Several works have established correlation between perceived task difficulty and mental workload [23, 34]. Mental effort is related to the characteristics of subjects, such as their driving experience. Since the mental effort varies over time and is

different among subjects, perceived task difficulty, as rated by an experienced therapist, was adopted as the ground truth for subjects’ mental workload. In the discussion section, we demonstrate a strong linear correlation between perceived task difficulty rating and task performance, which justified our method of acquiring subjects’ mental workload.

Observer rating data were logged based on a 0-9 continuous rating scale, where larger ratings indicated a higher intensity. At the end of an assignment, which usually took five minutes, the therapist was prompted to provide a summary rating for each of the categories on subjects’ overall states.

### 2.4.3 Subjects

This study was conducted with the approval from the Vanderbilt University Institutional Review Board. Twenty subjects (19 males, 1 female, mean age: 15.29 years) with a clinical diagnosis of ASD took part in the study. One subject had a driver’s license and three subjects had driver’s permits. Their ASD assessment results and IQs are reported in Table 2-1. Subjects attended six sessions on different days. During each visit, spanning approximately 60 minutes, they completed three preselected assignments. We measured their EEG response before the session for a three minutes baseline period and during the session.

Table 2-1. Subject Characteristics

	<b>Sample size</b>	<b>Mean</b>	<b>Standard deviation</b>
Chronological age in years	20	15.29	1.65
ADOS total raw score	16	13.56	3.67
ADOS severity score	16	7.81	1.42
SRS-2 total raw score	20	97.85	26.97
SRS-2 T-score	20	75.45	9.70
SCQ lifetime total score	19	20.84	9.48
IQ	15	108.93	17.47

ADOS = autism diagnostic observation schedule, SRS-2 = social responsiveness scale, second edition, SCQ = social communication questionnaire. Sample size varies due to missing data.

## 2.5 Methods

The continuous rating data were transformed into binary classes, low intensity class and high intensity class, for building models of affect and workload. As a first step we chose to develop binary classifiers. With more data and experience, multiclass classifiers will be developed in the future. The thresholds used

for the transformation were chosen by the therapist before conducting the driving experiment. For each category, if the rating score was less than the threshold, the corresponding assignment was labeled as low intensity class, otherwise it was labeled as high intensity class. The thresholds for engagement, enjoyment, boredom, frustration, and difficulty were 6, 6, 2, 2, and 5, respectively.

### **2.5.1 Signal processing**

Out of the 120 sessions (20 subjects  $\times$  6 visits), raw EEG data from 111 sessions were processed. Inevitable and unforeseen events, such as subjects selecting the wrong assignments, or the system being restarted due to eye tracker failure, led to the loss of some data. The spikes in EEG data were first removed by slew rate limiting. Then a 0.2-45 Hz bandpass filter was applied. Filtered data from baseline and each assignment were segmented into one-second epochs with 50% overlap. Corrupt epochs were identified and removed, and eye movement and muscular artifacts were corrected automatically. A detailed description of signal preprocessing procedure can be found in our previous paper [29].

For both baseline and assignment, EEG data that contained less than 60 artifact-free epochs were discarded. In the end, a total of 269 assignments were used for affect and workload recognition. The mean and standard deviation of the number of artifact-free epochs for those assignments were 296.42 and 150.40, respectively, whereas for the corresponding baseline EEG data, the number of useful epochs had a mean value of 170.12 and standard deviation of 52.61.

### **2.5.2 Feature generation**

A recent literature review on EEG features for emotion recognition quantitatively evaluated and compared different feature types [26]. The results showed that the first difference feature in statistical features, fractal dimension (FD) features, and higher order crossings (HOC)-based features performed well and were frequently selected by feature selection methods. Other feature types, such as higher order spectra (HOS) bicoherence features and Hilbert-Huang spectrum (HHS) features, contained valuable features as well. As a first step to explore EEG-based affect and workload recognition, statistical features, FD features, and HOC-based features were selected because of reported strong performances in other studies [24, 26, 35] and relatively simple implementation and less computational demand compared to HOS bicoherence and HHS features. Power features are the most popular features in EEG studies and therefore were included in the analysis.



We calculated five sets of features for EEG signals recorded from each electrode, summarized in Table 2-2. Time domain feature types captured the statistical measures (statistical features), signal complexity (FD features), and signal oscillation patterns (HOC-based features), whereas the frequency domain feature types characterized the strength of EEG oscillations at a given frequency range (bands and bins). Features extracted from artifact-free epochs belonging to the same baseline/assignment were averaged to obtain the feature vector for the corresponding baseline/assignment.

Table 2-2. Extracted Features

Features	Method/Parameter	No.
Statistics	mean $\mu_x$ , standard deviation $\sigma_x$ , first difference $\delta_x$ , standardized first difference $\tilde{\delta}_x$ , second difference $\gamma_x$ , standardized second difference $\tilde{\gamma}_x$ [35]	84
FD	Higuchi algorithm, $k_{\max} = 6$ [36]	14
HOC	backward difference operator, $k = 1, 2, \dots, 10$ [24]	140
Bands	Hanning tapering function, PSD (Welch's method), $\delta$ (1-4 Hz), $\theta$ (4-8 Hz), $\alpha$ (8-13 Hz), $\beta$ (13-30 Hz), $\gamma$ (30-44 Hz), normalization with log transformation	70
Bins	Hanning tapering function, PSD (Welch's method), 2-44 Hz ( $\Delta f = 2$ Hz), normalization with log transformation	294

### 2.5.3 Feature calibration

In order to train group-level models using EEG data collected from different subjects and different visits, it is necessary to remove feature variations resulting from time and individual differences. We developed a two-step feature calibration method, baseline feature subtraction followed by individualized feature normalization, to prepare the EEG features for group-level affect and workload recognition. The normalization step, as shown in Equation 2.1, rescaled the range of each subject's features based on the means of his/her features that belong to the low and high intensity classes. This requires each subject to provide data from both classes, e.g., examples from low engagement and high engagement. As a consequence, after individualized feature normalization, the number of examples reduced to 82 (6 subjects) for engagement, 184 (13 subjects) for enjoyment, 146 (10 subjects) for boredom, 248 (18 subjects) for frustration, and 244 (18 subjects) for workload. The effect of feature calibration is illustrated in Figure 2-2 using engagement examples. As can be seen, the examples are more separable after feature calibration.

$$f' = (f - \overline{f_{low}}) / (\overline{f_{high}} - \overline{f_{low}}) \quad (2.1)$$

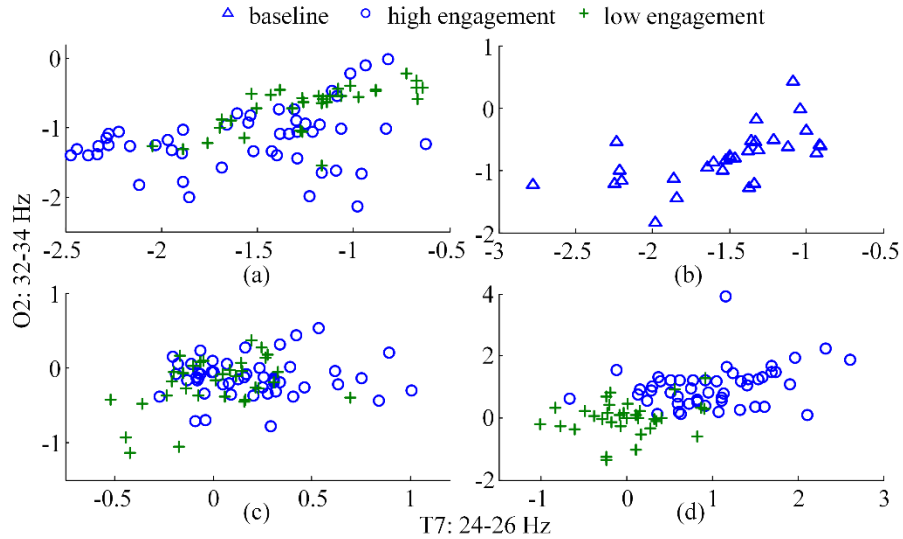


Figure 2-2. Comparison of 2D Feature Scatter Plot for Engagement (a) Before Feature Calibration, (b) Baseline Feature Distribution, (c) After Baseline Removal, and (d) After Feature Normalization

#### 2.5.4 Feature selection and classification

For the purpose of identifying discriminative features, we ranked the features based on univariate statistical tests. The one-way analysis of variance (ANOVA) test was used to rank features in descending order of the F-values. In other words, a larger F-value indicates the feature has greater discriminative power. Then, model training and evaluation was conducted on a subset of top-ranked discriminative features, where the number of features increased from 3 to 45 by iteratively adding 3 features based on F-value ranking. For FD features, the maximum number of features was 14.

The k-nearest neighbors (kNN) method was used for model training and evaluation because it performed the best in our preliminary study [29]. Three key hyper-parameters (16 combinations) were tuned, including the number of nearest neighbors (1, 3, 9, or 27), distance metric (Manhattan or Euclidean), and weighting scheme (uniform-weighted or distance-weighted).

Nested cross-validation (CV) was used for model selection. The outer loop was leave-one-subject-out CV whereas the inner loop was stratified ten-fold CV with randomization. The inner CV compared the classification performance of kNN models using each of the 16 combinations of hyper-parameters. The best performing model was then evaluated with the test set in the outer CV. The macro-averaged  $F_1$  score of two classes was used as the scoring function to compare hyper-parameters. For imbalanced datasets, the macro- and micro-averaged methods are more suitable than classification accuracy for representation of the results. However, in the case of binary classification, the micro-averaged method is the same as accuracy

measure. Therefore, we selected the macro-averaged method in this study. The classification result of the outer CV is the macro-averaged  $F_1$  score of the combined results of all the subjects. Since randomization was used in the inner CV, we performed 50 repetitions of nested CV to acquire more robust classification results. Standardization was used to preprocess the features.

## 2.6 Results

### 2.6.1 Classification results

Figure 2-3 summarizes the classification results of affect and workload recognition with respect to feature types and the number of features. For engagement recognition, power features from bins performed the best with 18 selected features. The performances of statistical features, HOC-based features, and power features from bands were less accurate (about 0.04). FD features scored significantly lower (by at least 0.2) than the other feature types. With only 3 features, power features from bands and bins reached a high macro-averaged  $F_1$  score of 0.90. In the case of enjoyment recognition, HOC-based features outperformed the other feature types significantly with a score of 0.88. The second best feature type was power features from bins with a score of 0.79, closely followed by statistical features and power features from bands. FD features performed the worst in enjoyment recognition as well. Similarly, HOC-based features achieved a much higher score for recognizing frustration. In terms of boredom recognition, power features from bins and HOC-based features performed the best with power features from bins performed slightly better. As far as workload recognition is concerned, the performances of HOC-based features, power features from bands and bins were comparable to each other.

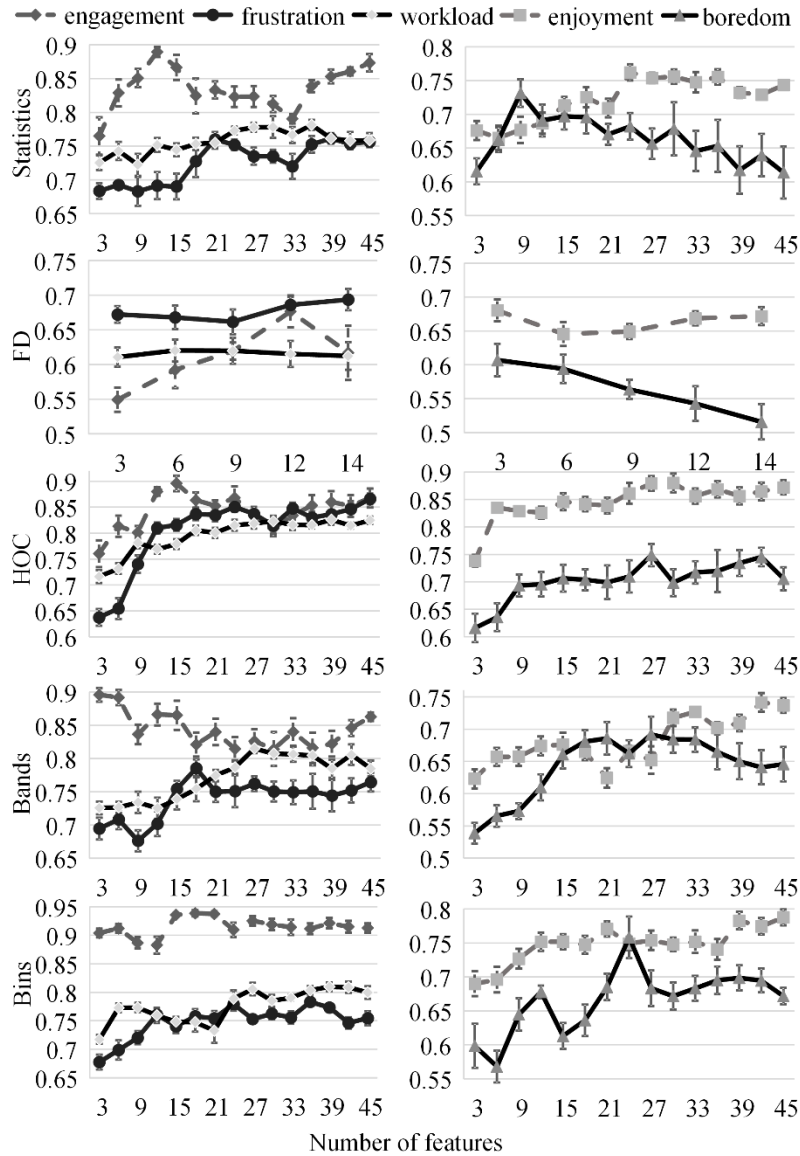


Figure 2-3. Classification Results Evaluated Based on Macro-averaged  $F_1$  Score

On average, except engagement recognition, HOC-based features were superior among the five feature types, especially in recognizing enjoyment and frustration. For engagement and boredom, power features from bins achieved the highest accuracy. The top feature types and the numbers of features selected by kNN were 18 power features from bins for engagement, 30 HOC-based features for enjoyment, 24 power features from bins for boredom, 45 HOC-based features for frustration, and 30 HOC-based features for workload. Given these feature types and the numbers of features, the most commonly selected hyper-parameters were 1 nearest neighbor, uniform weight, and Euclidean distance metric for engagement; 27

nearest neighbors, distance weight, and Manhattan distance metric for enjoyment; 3 nearest neighbors, uniform weight, and Euclidean distance metric for boredom; and 27 nearest neighbors, uniform weight, and Manhattan distance metric for frustration and workload. The final classification results using the best features with the best performing hyper-parameters are shown in Table 2-3. We list the precision, recall, and  $F_1$  scores of the leave-one-subject-out CV for both low and high intensity classes. High precision score, or positive predictive value, indicates that the classifier is returning accurate results. High recall score, or true positive rate, shows that the classifier is returning a majority of all positive results. These results imply that models based on EEG activation are able to detect with high accuracy subjects' states of low engagement, low enjoyment, high frustration, and high workload. Boredom recognition had relatively low accuracy. Because the number of examples in the low intensity class was four times larger than that of the high intensity class for boredom, the performance of boredom recognition might improve with more examples of high intensity boredom.

Table 2-3. Classification Results Using the Best Performing Features and Hyper-parameters

	<b>Intensity</b>	<b>Examples</b>	<b>Precision</b>	<b>Recall</b>	<b>F1 score</b>
Engagement	low	33	0.97	0.91	0.94
	high	49	0.94	0.98	0.96
Enjoyment	low	77	0.88	0.87	0.88
	high	107	0.91	0.92	0.91
Boredom	low	118	0.92	0.92	0.92
	high	28	0.64	0.64	0.64
Frustration	low	176	0.93	0.93	0.93
	high	72	0.82	0.82	0.82
Workload	low	122	0.83	0.89	0.86
	high	122	0.88	0.82	0.85

### 2.6.2 Discriminative features

To investigate which features and which electrodes provide the most information, we computed the relative frequency of each feature subtype and electrode. For each feature type and rating category, given the selected features that yield the best classification results, we counted the feature occurrence and electrode occurrence. Then, the occurrence counts were normalized by the total number of selected features and weighted by classification results. This step is important because it ensures that the relative frequencies

accounted for the performances of the discriminative features, and it enables comparison of discriminative features across rating categories. The results are shown in Figure 2-4. The feature usage is represented as histograms and the electrode usage is represented as topographies. The classification results, macro-averaged  $F_1$  scores, of the discriminative features are labeled as well.

Because backward models do not allow for a definitive physiologically-based interpretation [37], we did not attempt to link each of the identified features to the underlying neural sources. The discriminative features illustrated in Figure 2-4 were chosen jointly by the univariate feature selection method and kNN method to minimize the effect of noise on feature weights. For engagement, the majority of the discriminative features were located in the left hemisphere for power features from bands and bins, however, it was the opposite for HOC-based features. Power features were mostly selected from electrodes F7, F3, and T7, whereas HOC-based features were mostly drawn upon from electrodes FC6, F8, T8, and O2 with some features from electrodes F7, F3, and FC5. In terms of enjoyment, the important electrodes were AF3, FC5, FC6, F8, O2, and P8. In addition, electrodes F3 and O1 were frequently used by HOC-based features, and electrodes F7, AF4, and T7 were used often by statistical features. Power features from bands and bins were drawn upon from all the electrode locations. The discriminative features for boredom were mostly selected from electrodes FC5, F4, and O2. Additionally, electrode F7 was prominent for power features from bins and electrodes F8 and AF4 were important for HOC-based features. The prominent electrodes for frustration recognition were less clear. Locations O1, FC6, and F8 were mostly selected for statistical features. For HOC-based features, electrodes AF3, F7, F3, F4, P7, and O1 were frequently used. Power features from bands preferred electrodes F4 and F8, and power features from bins preferred electrodes T7, P8, and F4. In the case of workload recognition, for statistical features and power features the prominent electrodes were F7 and T8 with some importance given to F8 and P8. HOC-based features were mostly selected from electrodes AF4, F4, FC5, and T7. In general, power features from bins and bands were similar in electrode usage. However, power features from bins performed better than power features from bands. According to the feature usage histogram of power features from bins, features related to  $\beta$  and  $\gamma$  bands seem more valuable in recognizing affect and workload.

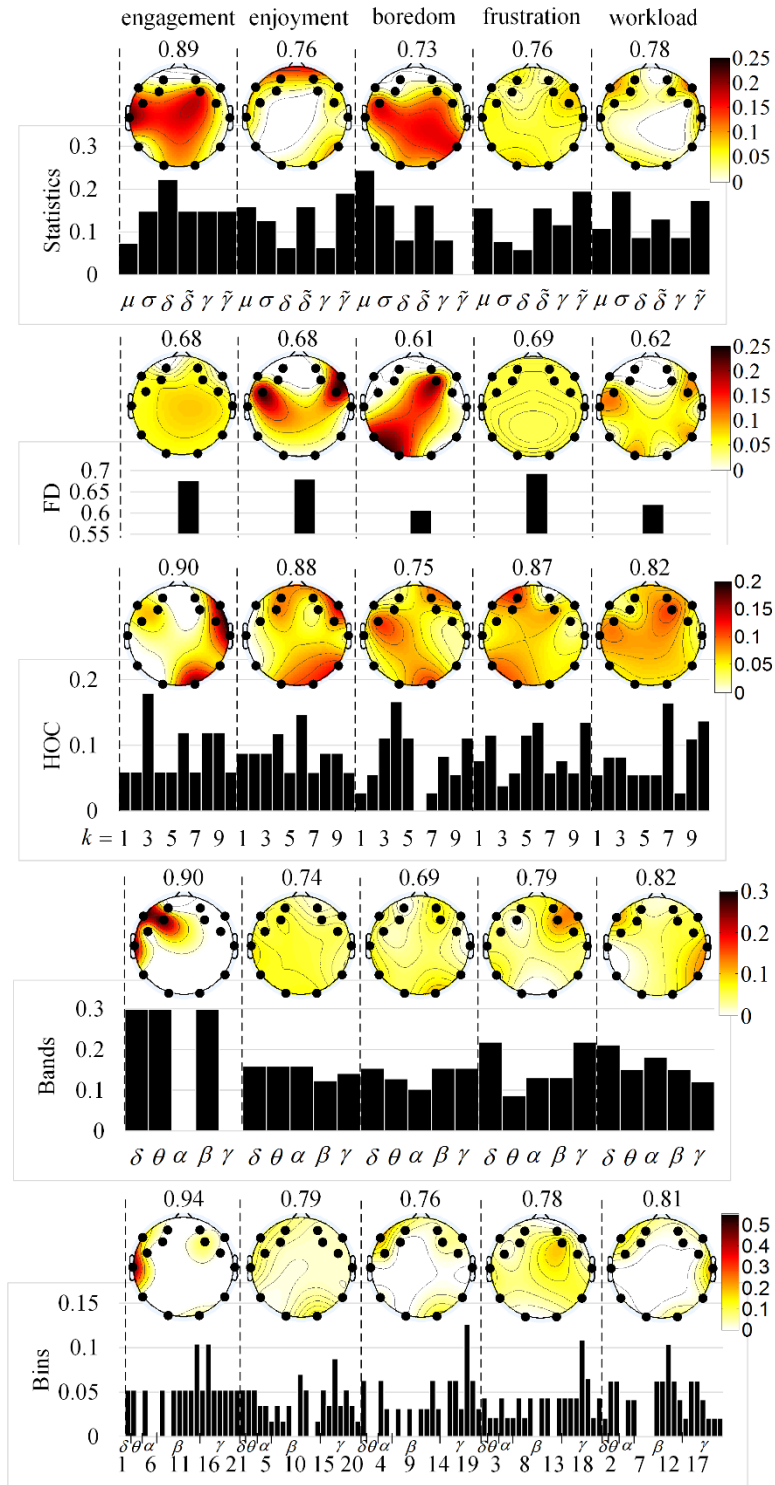


Figure 2-4. Feature Usage and Electrode Usage

## 2.7 Discussion

### 2.7.1 Validation of therapist's measures of mental states

The results presented in the previous section suggest that EEG-based affect and workload recognition is possible for adolescents with ASD and thus can be used to individualize driving training. However, the results need to be interpreted cautiously. It is unclear whether the overall rating data provided by a therapist can accurately capture subjects' affective states and mental workload. While expert rating is a widely used method [31] that we have used in this work, obtaining the ground truth of implicit user states is still a hard problem [38]. To examine the validity of the overall rating data and the thresholding values, we related the overall rating data to the performance data. The results are illustrated in Figure 2-5. The x axes are the performance data in terms of the number of errors that occurred in each assignment. The means and standard deviations of all the rating categories with respect to error count are shown as line plots. From the line plots, positive correlations between error count and three other variables: frustration, boredom, and difficulty, can be observed. We further evaluated the correlations based on the Pearson product-moment correlation coefficient. Strong positive correlations were found between frustration and error count ( $r = 0.57, p < 0.001$ ) and between difficulty and error count ( $r = 0.60, p < 0.001$ ). Moderate positive correlation between boredom and error count ( $r = 0.31, p < 0.001$ ) were found. In the cases of engagement ( $r = -0.29, p < 0.001$ ) and enjoyment ( $r = -0.25, p < 0.001$ ), weak negative correlations were found. Strong correlations between frustration and error count, and between difficulty and error count indicate that the overall rating data reflect the affective states and mental workload of individuals with ASD. Regarding the other three affective states, no conclusion could be drawn with respect to the validity of the overall rating data. In fact, their relationships with performance data are not entirely clear. In terms of the chosen thresholding values, it can be seen from the line plots in Figure 2-5 that the predetermined thresholding values approximately separate data into two camps, before 3 errors and after 4 errors. This observation is salient, especially for frustration and difficulty. In terms of boredom, the overall rating score decreased slightly when error count was 5 and 6, whereas for engagement and enjoyment the scores increased slightly when error count was 5 and 6. Overall, the choices of the thresholding values are consistent with performance data.



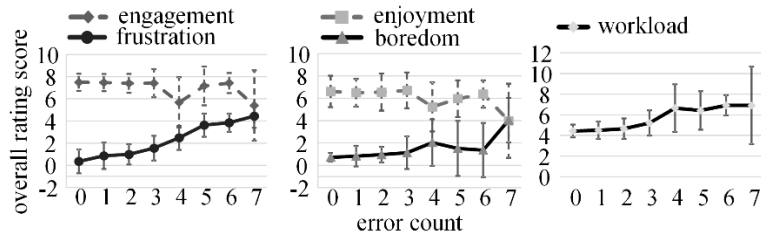


Figure 2-5. Overall Rating Data as a Function of Participants' Performance

Learning effect is one factor that would bias subjects' performance, workload, and affective states. We designed the experiment so that the first session and the last session consisted of the same assignments, one assignment from difficulty level two and two assignments from difficulty level five. Figure 2-6 shows the averaged error counts and therapist's ratings for session one and session six. Subjects' performance improved over the course of the six sessions due in part to the learning effect. For affect and workload recognition, we used the therapist's rating data as labels. This information is irrelevant to subjects' driving skills. As can be seen in Figure 2-6, subjects' frustration and workload levels were lower in session six than that in session one. Therefore, the learning effect biased the performance, but it should not bias the affective states and workload recognition. The learning effect is also one reason why we did not use the designed difficulty levels as labels for mental workload. In addition, we list subjects' error counts for all six levels of difficulty in Table 2-4. More difficult driving tasks did not necessarily indicate higher error rates. In fact, the correlation between levels of difficulty and error counts was  $r = -0.03$  ( $N = 360$ ). Because one attribute of mental workload, mental effort, varies due to learning effect and is different among subjects, it is not surprising that designed levels of difficulty does not correlate well with subjects' performance data.

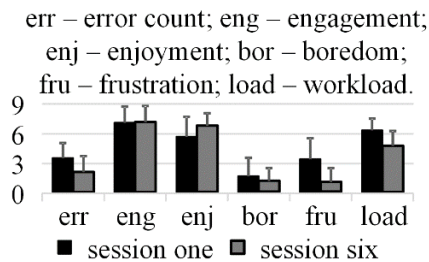


Figure 2-6. Learning Effect

Table 2-4. Error Counts

<b>Difficulty</b>	<b>Mean</b>	<b>STD</b>	<b>N</b>
level 1	2.62	1.79	58
level 2	2.84	1.94	43
level 3	2.47	1.64	60
level 4	2.05	1.57	60
level 5	2.39	1.60	79
level 6	2.70	1.67	60

### 2.7.2 Comparison with related works

Direct comparisons of classification accuracies between studies are difficult due to different experimental designs, subjects' characteristics, data preprocessing procedures, EEG recording devices, etc. The classification accuracies of a workload detector was more than 70% for a subset of subjects in [17]. Dijksterhuis et al. reported averaged accuracies up to 75-80% from lower EEG frequency ranges for workload classifications [19]. With respect to affective states recognition, [39] achieved an accuracy of 86.52% for distinguishing music likability. The best accuracy achieved in identifying emotional states was up to 77.78% over 5 subjects and 56.1% over 15 subjects in [40]. In another study, emotion recognition accuracy reached up to 83.33% for distinguishing 6 emotions and 100% for distinguishing fewer emotions [24]. Lan et al. [41] combined features with high intra-class correlation and improved accuracy to 73.1% for detecting positive negative emotions. Similarly, with combination of features, Liu et al. [42] achieved 87.02% accuracy in recognizing 2 emotions. Based on Table 2-3, our classification results scored 0.95 over 6 subjects for engagement, 0.895 over 13 subjects for enjoyment, 0.78 over 10 subjects for boredom, 0.875 over 18 subjects for frustration, and 0.855 over 18 subjects for workload. Overall, these results are comparable to the extant literature.

As far as the feature generation approaches are concerned, HOC-based features were shown to be superior in detecting emotional states. HOC-based features outperformed statistical features in [24] and power features from bands and bins in [26] for emotion recognition. Power features from bands and bins performed well for engagement and workload recognition. This is in line with other studies that used power features to monitor task engagement [20, 21], and analyzed correlations between power features and workload and engagement levels [23]. In addition, our results indicate that power features from bins are more valuable for engagement recognition compared to the rest of the feature types. Power features from bins outperformed power features from frequency bands in general. This trend is in accordance with [26].

The driving tasks were designed to resemble real world scenarios. The complexity of the task requires working memory and long-term memory, visuospatial processing, visual and auditory processing, attention and emotion regulation, and decision making. According to the electrode usage results, features derived from frontal electrodes were the most discriminative for affect and workload recognition. Temporal electrodes were also frequently used for engagement and workload recognition. Additionally, occipital electrodes were selected often for engagement, enjoyment, boredom, and frustration recognition. The discriminative features for enjoyment and frustration were mostly drawn upon from parietal electrodes as well. In terms of feature usage, features related to  $\beta$  and  $\gamma$  bands seem more valuable. These results are consistent with previous studies. Dijksterhuis et al. found that driver's workload classifications were most accurate when based on high frequencies and the frontal electrodes [19]. Frontal and parietal electrodes were found by Lin et al. [22] to be most informative for classifying emotional states. In [39], frontal, prefrontal, temporal, and occipital electrodes correlated significantly with ratings of music likability. Compared to other power features, features from  $\beta$  and  $\gamma$  bands were more discriminative in [39] and [26].

### 2.7.3 Limitations and future works

The main limitation of the current work comes from the small sample size. Unlike studies that could record large amount of samples per session, at most three samples were available per session in this study. Extracted EEG features were used to recognize the affective states and workload in each assignment. We did not attempt to detect the affective and workload changes based on epochs due to the requirement of human therapist's ground truth rating. It is not feasible to ask a therapist to provide rating data every few seconds. A secondary reason is that the individualized adaptation should not occur so quickly. For performance-based system adaptation, we do not reduce task difficulty whenever a driving error is detected. Similarly, for affect- and workload-based system adaptation, we will select the next driving task based on the overall states during the entire assignment instead of the states detected during the last few seconds of data.

There are several other limitations to address. First, a large portion of EEG data were removed in the process of artifact removal. It is worthwhile to explore how eye movements and muscular artifacts influence the classification results. They may improve the affective states and mental workload recognition [15, 19]. Second, different affective states may have an influence on mental workload recognition and vice versa. That is to say, in this study affective states are likely to co-vary with mental workload. Whether this confounding factor inflates the results, or could be harnessed to improve the classification accuracies,

requires further study. Third, the current work is limited to offline analysis. The future system needs to close the loop between the VR-based simulator and subjects using EEG signals to achieve individualized intervention. In addition, different strategies for VR-based driving system adaptation will be explored and the results will be subjected to comparison with performance-based system adaptation. Majority voting is one method to combine affect and workload prediction results for adaptive automation.

## **2.8 Conclusion**

We integrated an EEG input modality into a novel VR-based driving simulator which was developed for ASD intervention. EEG data as well as a therapist's overall rating data on five categories (engagement, enjoyment, boredom, frustration, and difficulty) were collected from 20 subjects diagnosed with ASD over a total of 120 sessions. Models of affect and workload were trained to provide the basis for a future EEG-based passive BCI system, which has the potential to tailor the driving skill training for specific individuals with ASD based on their affective states and mental workload. We systematically evaluated and compared five feature generation approaches with univariate feature selection method and the kNN algorithm. The classification results imply that models based on EEG activations are able to detect with high accuracy the states of low engagement, low enjoyment, high frustration, and high workload for ASD population. Boredom recognition had relatively low accuracy. In the end, classification models were built using power features from bins for engagement and boredom, and using HOC-based features for the rest of the states. The most discriminative features for affect and workload recognition were extracted from frontal electrodes. The analyses on EEG data collected from real world tasks with ASD population demonstrated the feasibility of EEG-based ASD intervention individualization and provided insight on the performance of several different feature types in this context. However, despite all its promise the current work is limited to binary classification and offline analysis after extensive artifact rejection. Thus while the current work is the first step towards an adaptive driving simulator for ASD intervention, the true potential of the developed models to measure the flow states of individuals with ASD based on online predictions of their affective states and mental workload requires further exploration in the future.

## REFERENCES

- [1] J. Baio, L. Wiggins, D. L. Christensen, M. J. Maenner, J. Daniels, Z. Warren, M. Kurzius-Spencer, W. Zahorodny, C. R. Rosenberg, and T. White, "Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014," *MMWR Surveillance Summaries*, vol. 67, p. 1, 2018.
- [2] C. Lord, E. H. Cook, B. L. Leventhal, and D. G. Amaral, "Autism spectrum disorders," *Neuron*, vol. 28, pp. 355-363, 2000.
- [3] S. M. Cox, D. J. Cox, M. J. Kofler, M. A. Moncrief, R. J. Johnson, A. E. Lambert, S. A. Cain, and R. E. Reeve, "Driving simulator performance in novice drivers with autism spectrum disorder: The role of executive functions and basic motor skills," *Journal of autism and developmental disorders*, vol. 46, pp. 1379-1391, 2016.
- [4] S. Classen, M. Monahan, and Y. Wang, "Driving characteristics of teens with attention deficit hyperactivity and autism spectrum disorder," *American journal of occupational therapy*, vol. 67, pp. 664-673, 2013.
- [5] B. P. Daly, E. G. Nicholls, K. E. Patrick, D. D. Brinckman, and M. T. Schultheis, "Driving behaviors in adults with Autism Spectrum Disorders," *Journal of autism and developmental disorders*, vol. 44, pp. 3119-3128, 2014.
- [6] B. Reimer, R. Fried, B. Mehler, G. Joshi, A. Bolfek, K. M. Godfrey, N. Zhao, R. Goldin, and J. Biederman, "Brief report: Examining driving behavior in young adults with high functioning autism spectrum disorders: A pilot study using a driving simulation paradigm," *Journal of autism and developmental disorders*, vol. 43, pp. 2211-2217, 2013.
- [7] J. Wade, D. Bian, L. Zhang, A. Swanson, M. Sarkar, Z. Warren, and N. Sarkar, "Design of a virtual reality driving environment to assess performance of teenagers with ASD," in *International Conference on Universal Access in Human-Computer Interaction*, 2014, pp. 466-474.
- [8] E. Sheppard, D. Ropar, G. Underwood, and E. van Loon, "Brief report: Driving hazard perception in autism," *Journal of autism and developmental disorders*, vol. 40, pp. 504-508, 2010.
- [9] N. B. Cox, R. E. Reeve, S. M. Cox, and D. J. Cox, "Brief Report: Driving and young adults with ASD: Parents' experiences," *Journal of autism and developmental disorders*, vol. 42, pp. 2257-2262, 2012.
- [10] A. C. Stahmer, L. Schreibman, and A. B. Cunningham, "Toward a technology of treatment individualization for young children with autism spectrum disorders," *Brain research*, vol. 1380, pp. 229-239, 2011.
- [11] M. R. Kandalaft, N. Didehbani, D. C. Krawczyk, T. T. Allen, and S. B. Chapman, "Virtual reality social cognition training for young adults with high-functioning autism," *Journal of autism and developmental disorders*, vol. 43, pp. 34-44, 2013.
- [12] E. Bekele, Z. Zheng, A. Swanson, J. Crittendon, Z. Warren, and N. Sarkar, "Understanding how adolescents with autism respond to facial expressions in virtual reality environments," *IEEE transactions on visualization and computer graphics*, vol. 19, pp. 711-720, 2013.
- [13] J. Wade, L. Zhang, D. Bian, J. Fan, A. Swanson, A. Weitlauf, M. Sarkar, Z. Warren, and N. Sarkar, "A gaze-contingent adaptive virtual reality driving environment for intervention in individuals with autism spectrum disorders," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 6, p. 3, 2016.
- [14] K. Kiili, "Digital game-based learning: Towards an experiential gaming model," *The Internet and higher education*, vol. 8, pp. 13-24, 2005.
- [15] A. Nijholt, D. P.-O. Bos, and B. Reuderink, "Turning shortcomings into challenges: Brain-computer interfaces for games," *Entertainment Computing*, vol. 1, pp. 85-94, 2009.
- [16] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*: Harper & Row, 1990.
- [17] J. Kohlmorgen, G. Dornhege, M. Braun, B. Blankertz, K.-R. Müller, G. Curio, K. Hagemann, A. Bruns, M. Schrauf, and W. Kincses, "Improving human performance in a real operating

- environment through real-time mental workload detection," *Toward Brain-Computer Interfacing*, pp. 409-422, 2007.
- [18] Y.-T. Wang, K.-C. Huang, C.-S. Wei, T.-Y. Huang, L.-W. Ko, C.-T. Lin, C.-K. Cheng, and T.-P. Jung, "Developing an EEG-based on-line closed-loop lapse detection and mitigation system," *Frontiers in neuroscience*, vol. 8, p. 321, 2014.
- [19] C. Dijksterhuis, D. de Waard, K. Brookhuis, B. Mulder, and R. de Jong, "Classifying visuomotor workload in a driving simulator using subject specific spatial brain patterns," *Frontiers in neuroscience*, vol. 7, p. 149, 2013.
- [20] F. G. Freeman, P. J. Mikulka, L. J. Prinzl, and M. W. Scerbo, "Evaluation of an adaptive automation system using three EEG indices with a visual tracking task," *Biological psychology*, vol. 50, pp. 61-76, 1999.
- [21] D. Szafir and B. Mutlu, "Pay attention!: designing adaptive agents that monitor and improve user engagement," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2012, pp. 11-20.
- [22] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "EEG-based emotion recognition in music listening," *Biomedical Engineering, IEEE Transactions on*, vol. 57, pp. 1798-1806, 2010.
- [23] C. Berka, D. J. Levendowski, M. N. Lumicao, A. Yau, G. Davis, V. T. Zivkovic, R. E. Olmstead, P. D. Tremoulet, and P. L. Craven, "EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks," *Aviation, space, and environmental medicine*, vol. 78, pp. B231-B244, 2007.
- [24] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion Recognition From EEG Using Higher Order Crossings," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14, pp. 186-197, 2010.
- [25] M.-K. Kim, M. Kim, E. Oh, and S.-P. Kim, "A review on the computational methods for emotional state estimation from the human EEG," *Computational and mathematical methods in medicine*, vol. 2013, 2013.
- [26] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *Affective Computing, IEEE Transactions on*, vol. 5, pp. 327-339, 2014.
- [27] L. M. Oberman, E. M. Hubbard, J. P. McCleery, E. L. Altschuler, V. S. Ramachandran, and J. A. Pineda, "EEG evidence for mirror neuron dysfunction in autism spectrum disorders," *Cognitive Brain Research*, vol. 24, pp. 190-198, 2005.
- [28] M. Murias, S. J. Webb, J. Greenson, and G. Dawson, "Resting state cortical connectivity reflected in EEG coherence in individuals with autism," *Biological psychiatry*, vol. 62, pp. 270-273, 2007.
- [29] J. Fan, J. W. Wade, D. Bian, A. P. Key, Z. E. Warren, L. C. Mion, and N. Sarkar, "A Step towards EEG-based brain computer interface for autism intervention," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, pp. 3767-3770.
- [30] H. H. Jasper, "The ten twenty electrode system of the international federation," *Electroencephalography and Clinical Neurophysiology*, vol. 10, pp. 371-375, 1958 1958.
- [31] R. S. Baker, S. K. D'Mello, M. M. T. Rodrigo, and A. C. Graesser, "Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments," *International Journal of Human-Computer Studies*, vol. 68, pp. 223-241, 2010.
- [32] N. Moray, *Mental workload: Its theory and measurement* vol. 8: Springer Science & Business Media, 2013.
- [33] B. Cain, "A review of the mental workload literature," Defence Research And Development Toronto (Canada)2007.
- [34] G. F. Wilson and C. A. Russell, "Performance enhancement in an uninhabited air vehicle task using psychophysiologicaly determined adaptive aiding," *Human factors*, vol. 49, pp. 1005-1018, 2007.

- [35] T. Y. Chai, S. S. Woo, M. Rizon, and C. S. Tan, "Classification of human emotions from EEG signals using statistical features and neural network," in *International*, 2010, pp. 1-6.
- [36] A. Accardo, M. Affinito, M. Carrozzi, and F. Bouquet, "Use of the fractal dimension for the analysis of electroencephalographic time series," *Biological cybernetics*, vol. 77, pp. 339-350, 1997.
- [37] S. Haufe, F. Meinecke, K. Görgen, S. Dähne, J.-D. Haynes, B. Blankertz, and F. Bießmann, "On the interpretation of weight vectors of linear models in multivariate neuroimaging," *Neuroimage*, vol. 87, pp. 96-110, 2014.
- [38] A.-M. Brouwer, T. O. Zander, J. B. van Erp, J. E. Korteling, and A. W. Bronkhorst, "Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls," *Frontiers in neuroscience*, vol. 9, p. 136, 2015.
- [39] S. K. Hadjidimitriou and L. J. Hadjileontiadis, "Toward an EEG-based recognition of music liking using time-frequency analysis," *Biomedical Engineering, IEEE Transactions on*, vol. 59, pp. 3498-3510, 2012.
- [40] A. T. Sohaib, S. Qureshi, J. Hagelbäck, O. Hilborn, and P. Jerčić, "Evaluating classifiers for emotion recognition using EEG," in *International Conference on Augmented Cognition*, 2013, pp. 492-501.
- [41] Z. Lan, O. Sourina, L. Wang, and Y. Liu, "Real-time EEG-based emotion monitoring using stable features," *The Visual Computer*, vol. 32, pp. 347-358, 2016.
- [42] Y. Liu and O. Sourina, "Real-time subject-dependent EEG-based emotion recognition algorithm," in *Transactions on Computational Science XXIII*, ed: Springer, 2014, pp. 199-223.

## CHAPTER 3

### A ROBOTIC COACH ARCHITECTURE (ROCARE)<sup>2</sup>

#### 3.1 Brief Summary

The population in the US is aging rapidly. Many older adults suffer from functional decline and/or cognitive impairment. Physical exercise, cognitive stimulation, and social engagement have been found to be beneficial for the physical and mental health of older adults with and without cognitive impairment. Multimodal strategies tailored to the individual appear most successful, but are resource intensive. These lead to the design and development of socially assistive robotic (SAR) systems to administer activity-oriented therapies. The objective of this work is to design a generalized model of interaction between a SAR and one to many older adults and to test the feasibility and older adults' acceptance on one-to-one and multi-user HRI.

The generalized model of interaction we designed is a multi-user engagement-based robotic coach system architecture (ROCARE). ROCARE is a user-centric model tied to core area of engagement and featured multi-user HRI and individualized activity management for long-term engagement. In addition, ROCARE incorporated both explicit and implicit states that can be detected by models of people. The feasibility of individualized activity management was tested by 11 older adults. We developed a semi-autonomous SAR system to administer five activities that were both passive and active. Prompts and reinforcements were developed and embedded in the system. A Kinect RGBD sensor detected in real time the gestures of the participants during the chair exercise activity. EEG and galvanic skin response (GSR) signals were recorded and aligned with each activity for offline analysis of activity engagement. The feasibility of multi-user HRI was tested by 14 older adults in the form of triadic interaction. A fully autonomous SAR system was developed to administer a gesture-based imitation game. Kinect was used to detect simultaneously two gestures performed by older adults. To test older adults' acceptance, we developed a robot acceptance scale (RAS) adapted from the unified theory of acceptance and use of technology framework.

For both one-to-one and triadic HRI, participants' perceptions as measured by RAS were more positive after the session. The engagement index computed from participants' EEG signals had strong

---

<sup>2</sup> © 2017 IEEE. Reprinted, with permission, from Jing Fan, Dayi Bian, Zhi Zheng, Linda Beuscher, Paul Newhouse, Lorraine Mion, and Nilanjan Sarkar, A robotic coach architecture for elder care (ROCARE) based on multi-user engagement models, *Neural Systems and Rehabilitation Engineering*, IEEE Transactions on, Aug. 2017.



correlation with their self-rating of activity preference, which indicates the potential for objectively measuring older adults' engagement intention and harnessing it to realize individualized activity management. Social communication between pairs of participants could be elicited by the robot as seen from both video recordings and head pose data. Collectively, these results suggest that i) ROCARE-based systems were well tolerated by the older adults and they were interested and engaged in robot-mediated activities; ii) our selection of the EEG feature has the potential for implementing individualized activity management; and iii) ROCARE-based interaction has the potential to involve more than one person and facilitate interpersonal communication.

There were 3 research papers published on this work.

- 1) **Fan, Jing**, Dayi Bian, Zhi Zheng, Linda Beuscher, Paul A. Newhouse, Lorraine C. Mion, and Nilanjan Sarkar. "A Robotic Coach Architecture for Elder Care (ROCARE) based on multi-user engagement models." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25, no. 8 (2017): 1153-1163.
- 2) Beuscher, Linda M., **Jing Fan**, Nilanjan Sarkar, Mary S. Dietrich, Paul A. Newhouse, Karen F. Miller, and Lorraine C. Mion. "Socially Assistive Robots: Measuring Older Adults' Perceptions." *Journal of gerontological nursing* 43, no. 12 (2017): 35-43.
- 3) **Fan, Jing**, Linda Beuscher, Paul A. Newhouse, Lorraine C. Mion, and Nilanjan Sarkar. "A robotic coach architecture for multi-user human-robot interaction (RAMU) with the elderly and cognitively impaired." In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pp. 445-450. IEEE, 2016.

### 3.2 Abstract

The aging population with its concomitant medical conditions, physical and cognitive impairments, at a time of strained resources, establishes the urgent need to explore advanced technologies that may enhance function and quality of life. Recently, robotic technology, especially socially assistive robotics has been investigated to address the physical, cognitive, and social needs of older adults. Most system to date have predominantly focused on one-on-one human robot interaction (HRI). In this paper, we present a multi-user engagement-based robotic coach system architecture (ROCARE). ROCARE is capable of administering both one-on-one and multi-user HRI, providing implicit and explicit channels of communication, and individualized activity management for long-term engagement. Two preliminary feasibility studies, a one-on-one interaction and a triadic interaction with two humans and a robot, were

conducted and the results indicated potential usefulness and acceptance by older adults, with and without cognitive impairment.

### 3.3 Introduction

In 2010, 13% of the US population was 65 years or older and this number is projected to double by 2030 with the oldest-old, those 85 years and older, growing at the fastest pace; this is the group most likely to have problems with physical functioning, functional decline, cognitive impairment, dementia, falls, and injury [1-4]. Up to 70% of older adults will develop significant disabilities and 35% will eventually reside in assisted living or enter a nursing home [5]. Health care costs for the behavioral consequences of these disorders are staggering [3, 6]. Thus, maintaining or improving physical and cognitive function, promoting communication and social interaction, and enhancing engagement are pivotal in geriatric care.

Nonpharmacologic interventions for these disorders such as physical activity, exercise, social interaction and engagement, cognitive stimulation, music, art therapy, reminiscence therapy, and caregiver intervention have had inconsistent results [7, 8] and can be resource intensive. Additionally, considering nursing shortage and high staff turnover in long term care settings, there is an urgent need for efficacious strategies that are tailored to the individuals within resource strained environments. Recently, socially assistive robotic (SAR) systems appear promising in addressing the physical, cognitive and/or social needs of older adults. A SAR system, unlike robotic wheelchair and exoskeleton, provides assistance and/or achieves measurable user progress through social interaction [9]. As compared to other interactive technologies, SAR has the advantage of embedding novel quantitative metrics, sensor-based non-invasive methodologies, incorporating physical movement into realistically embodied interactions, and meaningfully responding to pivotal aspects of human engagement and behavior, and thus has substantial promise for impacting function and engagement of older adults.

Earlier work of SAR systems with older adults [10, 11] primarily fall into two categories: companion robots, generally animal shaped, for social engagement [12], and service type robots supporting independent living, such as intelligent reminder etc. [13]. There is a growing interest in SAR systems that act as a coach or a guide to engage and encourage users through a series of therapeutic tasks for enhancing their physical or cognitive functions, as well as their health conditions. *We refer to such systems as robotic coach systems.* Several investigators have used the Wizard of Oz (WoZ) experimental paradigm [14] or open-loop robotic systems [15]. These systems are limited in their capacity for HRI, requiring remote human control for change of robot behaviors, and often times requiring sophisticated users.

More recently, closed-loop robotic systems have been developed. Commercial robots NAO [16, 17], RoboPhilo [18], and Manoi-PF01 [19] were used to build robotic coach systems to assist older adults in performing physical exercises. Fasola and Matarić [20] designed and implemented a robotic coach system, Bandit, that monitored and encouraged older adults to perform chair exercises. Bandit personalized its interaction via task performance, progress, and session history. Tapus et al. [21] tested the effectiveness of robot-mediated cognitive intervention with dementia patients once per week over eight months and observed an improvement in task performance. These works relied on explicit task performance as feedback to adapt robot's behaviors. McColl et al. included implicit channel of communication in their robotic system Brian 2.1 [22]. Brian 2.1 was developed to engage older adults in eating activity and a cognitive stimulation activity, and had the capability of adapting its behavior based on the state of the activities as well as user's body language (attentive or distracted). These researchers also developed a robotic system Tangy [23, 24] for use in long-term care facilities to provide telepresence and group-based cognitive intervention. Robotic coach systems were also developed for stroke rehabilitation, autism intervention, and weight loss [21, 25, 26].

Most systems to date have predominantly focused on one-on-one interaction. Multi-user interaction is pivotal for fostering social interaction. Only two studies to our knowledge have investigated group-based closed-loop robot-mediated interaction for older adults. Kanoh and associates devised a robot-assisted activity program comprised of one robot with five to six human participants, but required one human assistant to mediate communication between the robot and the participants [27]. Louie and associates were able to provide autonomous interaction by the robot with an individual, but not between individuals [24]. The objective of this work was to develop a robotic coach system architecture that allows effective interaction with one or multiple older adults and achieve long-term engagement for the purpose of maintaining functional abilities as well as socialization. In this paper, we present the mathematical models of the system architecture which a) is capable of one-on-one interaction and multi-user interaction; b) contains both explicit and implicit channels of communication; and c) allows dynamic adaptive robotic behavior and activity management based on real-time human interaction. Further, we performed two feasibility studies on older adults to assess our design paradigm and test on older adults' acceptance of the robotic coach system.

The paper is organized as follows. Section 3.4 presents the mathematical models of the RObotic Coach ARchitecture for Elder care (ROCARE) and places it in context with existing SAR architectures. Section 3.5 describes the preliminary feasibility studies and system implementation. Section 3.6 and 3.7 presents the results and discusses their implications, respectively. Finally, we summarize the contributions of the paper and highlight future directions in section 3.8.

### 3.4 Design of a Robotic Coach System Architecture

The primary goals of the ROCARE are: a) it should be able to adapt its behavior to each individual user as the HRI progresses; b) it needs to perform quantitative measurements of the user's task performance on activities, as well as the user's affective states and gaze position; c) the robotic coach system must be designed so that it can be operated by a non-technical caregiver; and d) the system can target activities designed to be beneficial in addressing mobility and functioning simultaneously satisfying user's preferences and ability. To achieve these goals, ROCARE needs to have the following features: a) multimodal HRI; b) individualized robot behavior adaption; c) rigorous measurements and well-structured task design; and d) administrator friendly control panel.

ROCARE (Figure 3-1), which possesses the aforementioned characteristics, is comprised of five modules: *Sensing*, *Actuation*, *Database*, *Supervisory Controller* and *Graphical User Interface (GUI)*. A human administrator is responsible for initializing the session and monitoring task progression via the GUI. *Database* maintains a knowledge base of each user to facilitate the decision-making process of the *Supervisory Controller*. *Supervisory Controller* is the core element of ROCARE. It estimates the states of HRI and human-human interaction (HHI) based on engagement models, and generates control policies for dynamic system adaptation. Users interact with ROCARE through *Sensing* and *Actuation*. *Sensing* collects both implicit and explicit interaction cues from users, whereas *Actuation* performs the actions the system needs to take as determined by the *Supervisory Controller*. The five primary modules are composed of submodules that are responsible for specific functionalities.

#### 3.4.1 Comparison with existing SAR architectures

There are several existing SAR architectures designed for behavior intervention for children with autism spectrum disorder [28], functional intervention or companion purpose for older adults [13, 20, 24, 29-32], as well as other applications [33, 34]. All these architectures including ours have component(s) or module(s) dedicated to sensing and actuation, which provide the interface between SAR systems and targeted users; and decision making, for system behavior adaptation. SAR Architectures in [28] and [20] incorporated a database to store HRI history. In ROCARE, the submodule *interaction memory* in *Database* serves a similar purpose.

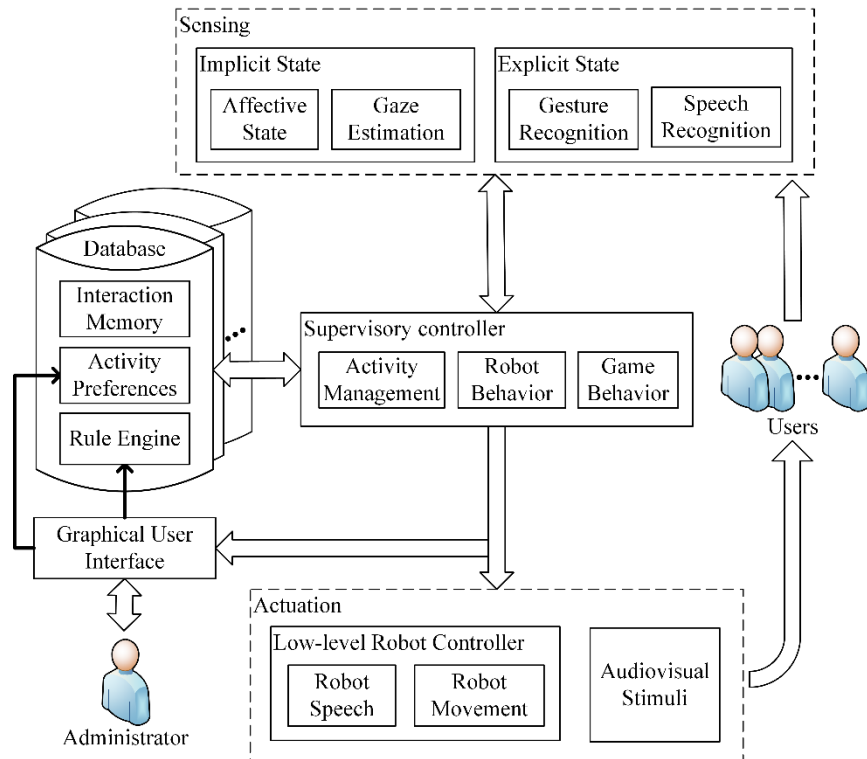


Figure 3-1. Robotic Coach System Architecture (ROCARE)

Our approach is tightly coupled with the primary goals described earlier and results in some differences and rearrangements of the modules. First, in *Sensing* we modified and extended the Sensory Input Recognition and Analysis Modules proposed by Chan and Nejat [29, 30] and the User State and Activity State Modules presented by Louie et al. [24]. The *implicit state* submodule in ROCARE is dedicated to implicit channel of communication between the SAR system and users, whereas the *explicit state* submodule is dedicated to explicit channel of communication. Second, two submodules, *activity preferences* in *Database* and *activity management* in *Supervisory Controller*, were added to keep track of user’s preferences and select appropriate activities for the purpose of promoting engagement while maximizing the efficiency of the interaction. Third, we integrated a GUI to allow intuitive control, operation, and monitoring of the robotic coach system by an administrator.

Several characteristics distinguish ROCARE: a) mathematical models for each module and relationships among modules instead of simple interconnection; b) engagement models to capture the dynamics of HHI and HRI; c) capacity for both one-on-one interaction and multi-user interaction; and d) generalizability of the architecture for different HRI scenarios. In what follows, we describe each module along with its submodules in detail.

### 3.4.2 Engagement models

In HRI, engagement is a critical component, defined as the act of being occupied or involved in an external stimulus [35]. We capture the dynamics of HHI and HRI using engagement models. Our models leverages the models for multiparty engagement proposed by Bohus and Horvitz [36]. We adopted their idea of representing user engagement using three engagement variables but modified the model for each engagement variable. The three engagement variables for each agent  $a \in \{user(s), robot\}$  and interaction  $i \in \{HRI, HHI\}$  are: the engagement state  $ES_a^i(t)$ , the engagement action  $EA_a^i(t)$ , and the engagement intention  $EI_a^i(t)$ .

The engagement state  $ES_a^i(t)$  represents whether agent  $a$  is involved in interaction  $i$  and is modeled by a timed automaton with two states: *engaged* or *not-engaged* (Figure 3-2). We presume that all agents are in the state *engaged* at the beginning of the interaction. Since engagement is a collaborative process, agent  $a$  is *engaged* either with an engagement action  $EA_a^i(t)$  initiated by agent  $a$ , such as gestures or direct responses, or with engagement intention  $EI_a^i(t)$ , which indicates agent  $a$  is paying attention to other agents. Agent  $a$  becomes disengaged, in state *not-engaged*, if he/she is not actively involved in the interaction for *time\_out* amount of time. The engagement action  $EA_a^i(t)$  is estimated by a conditional statistical model of the form:

$$P(EA_a^i(t) | \Psi_a(t), \{ES_a^i(t-1)\}_{a \in \Omega_i}, \{ES_a^i(t)\}_{a \in \Omega_i}, \Lambda(t)) \quad (3.1)$$

Occurrence of engagement action of agent  $a$  in the interaction  $i$  depends on gestures, speech or direct inputs detected by the system, i.e., *explicit state* of agent  $a$  ( $\Psi_a(t)$ ); previous engagement states of all the agents in the interaction  $i$  ( $\{ES_a^i(t-1)\}_{a \in \Omega_i}$ ); current engagement states of all the agents in the interaction ( $\{ES_a^i(t)\}_{a \in \Omega_i}$ ); and the current game behavior ( $\Lambda(t)$ ). Similarly, the engagement intention  $EI_a^i(t)$  is estimated by:

$$P(EI_a^i(t) | \Gamma_a(t), \{ES_a^i(t-1)\}_{a \in \Omega_i}, \{ES_a^i(t)\}_{a \in \Omega_i}) \quad (3.2)$$

where  $\Gamma_a(t)$  denotes the agent's *implicit state* detected by the system, including affective states (engaged, bored, frustrated, etc.) as well as direction of attention measured by gaze position. We describe

$\Psi_a(t)$  and  $\Gamma_a(t)$  in more detail in the next section. For agent  $a = robot$ ,  $\forall t \in \mathbb{R}$ ,  $i \in \{HRI, HHI\}$ ,  $ES_a^i(t) = engaged$ ,  $EI_a^i(t) = true$ .

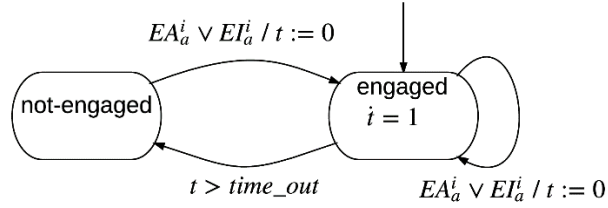


Figure 3-2. Timed Automaton Model of  $ES_a^i(t)$

### 3.4.3 Sensing and actuation

The multimodal HRI feature is reflected in *Sensing* and *Actuation*. *Sensing* is responsible for logging and interpreting data collected by sensors and cameras. It is composed of *implicit state* ( $\Gamma_a(t)$ ) and *explicit state* ( $\Psi_a(t)$ ). *Implicit state* facilitates the inference of the engagement intention of agent  $a$  through affective states recognition and gaze estimation. For example, in a multi-user interaction scenario, when the robot is interacting with one user, another user may be engaged by having eye contact with the robot even though he/she is not directly involved in the interaction. *Explicit state* aids the inference of the engagement actions. According to the context of the interaction, i.e., game behavior ( $\Lambda(t)$ ), detected gesture or speech inputs are engagement actions if they are directly related to task performance. Otherwise, based on the previous and current engagement states of all the agents, detected  $\Psi_a(t)$  may be social cues during interaction (an engagement action) or random noise (not an engagement action).

Both  $\Gamma_a(t)$  and  $\Psi_a(t)$  are detected by *Sensing* and are sent to the *Supervisory Controller* to estimate  $EA_a^i(t)$  and  $EI_a^i(t)$ . *Sensing* communicates with *Supervisory Controller* in two modes: a) sending current *implicit state* and *explicit state* upon request. For instance, the *Supervisory Controller* queries *gesture recognition* about the user's performance on the exercise motions during the physical exercise task. b) Whenever a significant event is detected. Example of a significant event is the user's gaze shifts away when the robot is dancing, which indicates the user is not interested or distracted; in this event, actions need to be taken to reengage the user. Behaviors of the robotic system are generated via *Actuation*, which consists of the *low-level robot controller* and *audiovisual stimuli*. *Low-level robot controller* manages and controls

the robot hardware to realize *robot behavior*, a.k.a. , while the *audiovisual stimuli* operates all the other hardware involved in the interaction, e.g. monitors, and updates the *game behavior*  $\Lambda(t)$ .

### 3.4.4 Database and graphical user interface

*Database* contains three main submodules: *interaction memory*, *activity preferences*, and *rule engine*.

It is individualized; in other words, each user  $a \in (\Omega - \{\text{robot}\})$  has his/her own database which is independent from other databases. *Interaction memory* stores the history of HHI and HRI, represented by the following tuple:

$$\langle ES_a^i(t), EA_a^i(t), EI_a^i(t), \Psi_a(t), \Gamma_a(t), \Lambda(t) \rangle_{a,i \in \Omega, t \in \mathbb{R}} \quad (3.3)$$

*Activity preferences* and *rule engine* submodules provide key information for activity management. They maintain three sets of parameters, including each user's degree of likes and dislikes regarding different types of activities  $AT$  ( $AP_a^{AT}$ ), the importance of each activity type for each user ( $AI_a^{AT}$ ), and the appropriate difficulty level of each activity type for each user ( $D_a^{AT}$ ). These parameters are updated during the interaction based on the following model:

$$AP_a^{AT} = f_1\left(\{ES_a^i(t)\}_{a,i \in AT, t \in \mathbb{R}}, f_{GUI}(AP_{GUI})\right) \quad (3.4)$$

$$AI_a^{AT} = f_2\left(\{\Lambda(t)\}_{t \in \mathbb{R}}, f_{GUI}(AI_{GUI})\right) \quad (3.5)$$

$$D_a^{AT} = f_3\left(\{\Psi_a(t), \Gamma_a(t), \Lambda(t)\}_{a,t \in \mathbb{R}}, f_{GUI}(D_{GUI})\right) \quad (3.6)$$

$AP_a^{AT}$  is a function of user's engagement state corresponding to activity  $AT$ , as well as direct input by an administrator through the GUI.  $AI_a^{AT}$  is a function of the history of the game behavior, i.e., past activities, and changes made via the GUI. And the difficulty level parameters  $D_a^{AT}$  is determined by the user's task performance and implicit state, as well as GUI inputs.  $f_{GUI}$  is a monotonically increasing function which weighs the direct inputs from the GUI. The sole purpose of the GUI is to allow non-experts to operate ROCARE and monitor the progress.



For illustrative purposes, we give an example of updating  $AP_a^{AT}$ . We assume that there are  $m$  types of activities and  $\sum_{AT=1}^m AP_a^{AT} = 1$ ; the user likes the activity if he/she is engaged for more than half of the activity duration and the weights for engagement and GUI inputs are  $w_{ES}$  and  $w_{GUI}$  respectively. Initially,  $\forall AT, AP_a^{AT} = 1/m$ . The robotic system starts with the activity “dance to music”. At the end of this activity, the new set of  $AP_a^{AT}$  is calculated as:

if  $AT = music$ ,

$$\Delta AP_a^{AT} = w_{ES} \cdot \left( \frac{\sum_{t=ATstart}^{ATend} ES_a^i(t)}{\sum_{t=ATstart}^{ATend} t - 0.5} \right) + w_{GUI} \cdot \Delta AP_{GUI} \quad (3.7)$$

$$AP_a^{AT} = \frac{1}{m} + \Delta AP_a^{AT} \quad (3.8)$$

$\forall AT \neq music$ ,

$$AP_a^{AT} = \frac{1}{m} - \frac{\Delta AP_a^{music}}{m-1} \quad (3.9)$$

### 3.4.5 Supervisory controller

*Supervisory Controller* is responsible for making autonomous decisions related to the robot’s behavior ( $\{EA_a^i(t)\}_{a=robot,i}$ ) and task adjustment ( $\Lambda(t)$ ) (e.g., repeat the exercise, establish mutual gaze, etc.) regarding the ongoing activity chosen by *activity management*. *Activity management* is dedicated to activity selection and scheduling by analyzing  $AP_a^{AT}$ ,  $AI_a^{AT}$ , and the goal of the system ( $G$ ). In a one-on-one interaction scenario, it can be realized by simply selecting the activity that has the highest  $AP_a^{AT} + AI_a^{AT}$  value. System goal can be represented by the engagement variables. For example, in one-on-one interaction scenarios, the system goal could be to maximize user engagement during HRI, i.e.,  $G = \max_{HRI}(ES)$ , whereas for the multi-user case, the goal could be to maximize user engagement with another user, i.e.,  $G = \max_{HHI}(ES)$ . *Robot behavior* and *game behavior* are either controlled by a reactive model or through learning algorithms. The control policies for *robot behavior* and *game behavior* are conditioned on task difficulty level  $D_a^{AT}$ , the engagement variables of all the users  $EV$ , previous robot behavior and game behavior  $H$ , and the goal of the system  $G$ .

$$\mathbf{EV} = \left\langle ES_a^i(t), EA_a^i(t), EI_a^i(t) \right\rangle_{a \in (\Omega - \{\text{robot}\}), i \in AT, t \in \mathbb{R}} \quad (3.10)$$

$$\mathbf{H} = \left\langle \left\{ EA_a^i(t') \right\}_{a=\text{robot}, i \in AT, t' < t}, \left\{ \Lambda(t') \right\}_{t' < t} \right\rangle \quad (3.11)$$

$$\pi_{\text{robot, game}} \left( \left\{ D_a^{AT} \right\}_{a \in (\Omega - \{\text{robot}\}), AT}, \mathbf{EV}, \mathbf{H}, \mathbf{G} \right) \quad (3.12)$$

### 3.5 Feasibility Studies

We conducted two preliminary feasibility studies to determine whether ROCARE a) can be used for both one-on-one and multi-user interaction; and b) could engage older adults' interest and participation. While the architecture does not assume any particular robot, our system was built around the NAO robot platform ([www.aldebaran.com](http://www.aldebaran.com)) because of its availability as well as its open architecture that allows relatively easy pathways for custom software development and integration with other devices.

#### 3.5.1 HRI scenarios

*Scenario 1 – individual user performing multiple activities.* This scenario was designed to explore older adults' behaviors and responses to ROCARE, and the feasibility of inferring their engagement intention variables ( $EI_a^i(t)$ ) based on implicit communication cues, in this specific case, through electrophysiological signals and gaze position ( $\Gamma_a(t)$ ). Several activities were selected: an orientation activity where the robot points to pictures hanging in the experiment room, simple math, observing the robot dance to music, a form of the “21 questions” game where the robot guesses the person's birth state, and joint chair exercises. Each participant sat in a straight back chair directly in front of and six feet away from the robot. Three pictures were hung on the walls of the experiment room at different locations. Electrophysiological sensors were placed on the head and the body of the participant. A Kinect RGBD sensor was used to augment NAO's vision for gesture recognition and gaze estimation. A researcher initiated and observed the session via the GUI and a one-way mirror in an adjacent room.

*Scenario 2 – paired users performing single activity.* We extended the interaction to allow simultaneous interaction with two older adults mediated by the robotic coach. The triadic HRI scenario consisted of introduction and “Simon says” game [37], where each individual and the robot took turns as Simon. Physiological sensors were excluded in this scenario.

### 3.5.1.1 One-on-one interaction

The implemented *Sensing* module is capable of electrophysiological signal collection, gaze estimation, gesture recognition, and speech recognition. We used a 14-channel Emotiv EPOC neuroheadset (www.emotiv.com) to record electroencephalography (EEG) signals, and a Biopac MP150 physiological data acquisition system (www.biopac.com) to collect physiological data. The bandwidth of the EEG signals is from 0.2 to 45Hz and the sampling rate is 128Hz. Regarding physiological signal, tonic and phasic responses from galvanic skin response (GSR), were logged with a sampling rate of 1000Hz. These signals have been shown to be sensitive to affective states in our previous work as well as others [38-40]. The signals were collected for offline analysis.

*Gaze estimation* was approximated by participant's head pose around yaw axis (horizontal head turn) extracted from the Kinect Face Tracking engine. For *gesture recognition*, we adapted a rule-based finite state machine (FSM) gesture recognition method [26] based on the upper body skeletal data from Kinect to accommodate for motor control declines in older adults. Both skeleton and head pose were updated at a frame rate of 30Hz. *Gesture recognition* was used during the joint chair exercise activity for monitoring participant's performance on an exercise motion demonstrated by the robot. Four gestures were recognized, including raise one arm up, raise both arms up, extend arms to the sides, and wave. The robot provided feedback prompts based on older adult's performance and then demonstrated the next exercise motion. *Speech recognition* was designed to understand three types of user responses: affirmative answer (e.g., yes, sure, ok, correct), negative answer (e.g., no, wrong), and repeat question (e.g., repeat). We are aware of the limitations of NAO's speech recognition software. It requires participants to speak loudly and is sensitive to different accents. Even though we informed the participants to speak loudly and clearly, and provided a word list that robot could understand, there were times when they forgot robot was not as intelligent as a human and would engage in conversation! For these experiments, we increased the robustness of *speech recognition* by asking the administrator to select the correct user response using the GUI. Both *gesture recognition* and *speech recognition* were in idle states unless invoked by the *Supervisory Controller*.

Figure 3-3(a) illustrates the *Supervisory Controller* for one-on-one interaction. *Activity management* scheduled the five activities in a predefined order as shown in the figure. The control policy for *robot behavior* and *game behavior* was modeled using hierarchical FSMs. We expanded the hierarchical FSM and participant's engagement model during the math activity, shown in Figure 3-3(a). The set of states in the hierarchical FSM model of control policy represents *robot behavior* and *game behavior*. The machine starts in state  $s_0$ , which has a refinement that is another FSM with states and transitions designed to realize the math activity. When the engagement state of the participant is *not-engaged*, the machine transitions to state  $s_1$ . In this state, robot gently prompts the participant to come back to the activity. The transition from

$s1$  to  $s0$  is guarded by *engaged* and is a history transition. When this transition is taken, the destination refinement  $s0$  resumes in whatever state it was last in. This ensures that the robot does not restart the activity from the beginning. The refinement FSM initially enters the state *Start activity* and sends the start marker to the electrophysiological data acquisition systems. For each math question, the robot either repeats the question or gives feedback based on the participant's response. The activity ends after finishing all the math questions or if the participant indicates his/her unwillingness to proceed. In this activity, the engagement action variable of the participant  $\{EA_a^i(t)\}_{a=user,i}$  was conditioned on *speech recognition* and *game behavior*, and the engagement intention variable  $\{EI_a^i(t)\}_{a=user,i}$  was conditioned on *gaze estimation*. Since *speech recognition* was enabled only when robot expected a response from the participant,  $\{EA_a^i(t)\}_{a=user,i}$  is true when *speech recognition* return values and is false otherwise. The participant's gaze, on the other hand, was monitored continuously. The Kinect Face Tracking engine tracked the head pose yaw angle from -45 degrees (turn towards the right) to 45 degrees (turn towards the left).  $\{EI_a^i(t)\}_{a=user,i}$  is true when the participant's gaze focuses on the robot, defined by  $|yawAngle| \leq 28$ . The engagement state transitions in Figure 3-3(a) indicates that when the participant looks away from the robot over a consecutive three-second time window, his/her engagement state  $\{ES_a^i(t)\}_{a=user,i}$  changes to not-engaged. The engagement state model omitted  $\{EA_a^i(t)\}_{a=user,i}$  because responding to the robot is usually accompanied by mutual gaze.

The robot's movement and speech were controlled through the NAOqi programming framework. A library of primitive robot motions, such as cheers, pointing, etc., were established. The primitive robot motions together with robot speech were building blocks for robot behaviors.

### 3.5.1.2 Triadic interaction

The chair exercise activity was expanded into a form of "Simon says" game in this HRI scenario. One player takes the role of Simon and instructs other players to perform physical movement. The other players should only follow the instructions prefaced with the phrase "Simon says". Due to the technical challenge of recognizing speech inputs from two individuals at the same time, we used a Razer Hydra (sixense.com), which has two separate controllers, to record trigger buttons click inputs instead. The control policy modeled by a hierarchical FSM is shown in Figure 3-3(b). There are three AND states, *Main Procedure*, *Gesture Checker*, and *User Input*, implemented using threads and processes with socket communication. The refinements of *Gesture Checker* and *User Input* control the communication between *Supervisory*

*Controller and Sensing.* Robot behavior and game behavior were defined in the refinement of *Main Procedure*. The initial state is *Meet each other*, when the robot and the participants introduce themselves and say “hi” to each other. In the next state the robot explains the rules to the participants. The transition from *Explain game rules* to *Robot plays Simon* takes place if both participants indicate understanding of the rules. The robot leads the chair exercise first. It checks the performance of each participant and provides feedback prompts. When the robot finishes three or four commands, each participant takes turns to play Simon. If the participant who plays Simon signals “Simon says” to the robot by pressing the trigger button, the robot mirrors his/her movement.

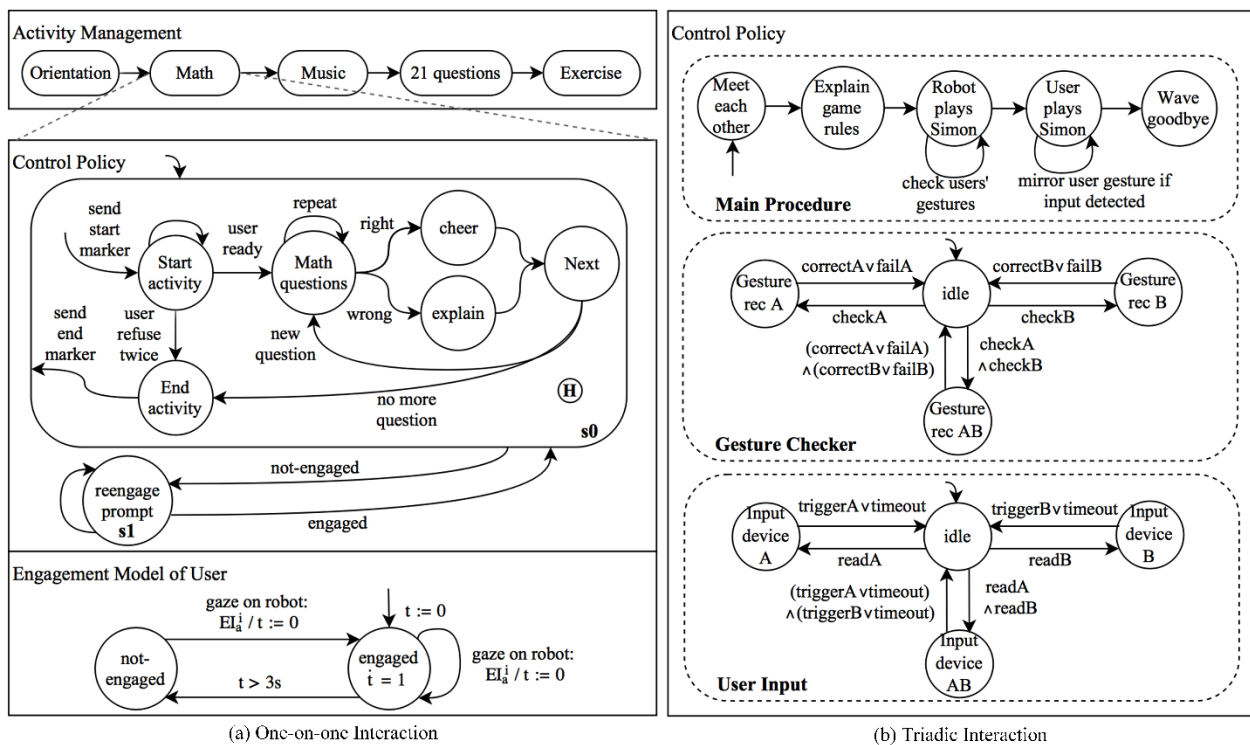


Figure 3-3. (a) Supervisory Controller Module for One-on-One Interaction (b) Control Policy for Triadic Interaction

### 3.5.2 Participants and protocol

Informed consents were obtained before the experiments, according to the protocol approved by Vanderbilt University Institutional Review Board.

*Scenario 1 – One-on-one interaction.* We recruited 11 community-residing older adults (6 females, 5 males, age: 66-94 years, mean: 82.5) of which 4 had a preexisting diagnosis of MCI or dementia. The entire session, approximately 60 minutes in duration, was video-recorded. Electrophysiological signals were

collected for a three-minute resting baseline and during HRI. A survey (Robot Acceptance Scale-RAS) was conducted pre- and post-experiment to determine the participants' acceptance and anticipated use of the robot on a 7-point scale (1 most positive to 7 most negative response). This survey was adapted from the Unified Theory of Acceptance and Use of Technology (UTAUT) and reflects the user's acceptance and intention to use new technology based on performance expectancy, effort expectancy, attitude toward using technology, and self-efficacy. The UTAUT framework posits that the person's pre-use attitudes influence the person's acceptance and use of the technology [41]. For this study, items were modified to reflect adults' interactions specific to robots. The final RAS consisted of 29 items (9 performance expectancy, 5 effort expectancy, and 15 attitude). Participants also completed a post-experiment questionnaire that provided opinions about the activities (from "extremely interesting (1)" to "extremely boring (7)").

*Scenario 2 – Triadic interaction.* We recruited 14 older adult participants (9 females, 5 males, age: 70-90 years, mean: 82.7) who were paired for simultaneous interaction with the robot. One pair of the participants had a formal diagnosis of MCI or dementia. Paired participants came to the lab once for approximately 30 minutes. EEG signals and the RAS were collected following the same procedure as in scenario 1. Participants also completed a pre- and post-experiment questionnaire on the degree to which they enjoyed interacting with and helping others. After the experiment, a questionnaire was provided to gain feedback on the level of enjoyment or interest with the activity.

## 3.6 Results

### 3.6.1 Data analysis methods

The Wilcoxon signed-rank test was applied to determine the survey's ability to be sensitive to change. RAS and its three subscales (performance expectancy, effort expectancy, and attitude) were subject to pre and post-experiment comparison. The Wilcoxon signed-rank test, a non-parametric statistical hypothesis test of median, was used because it does not assume normal distribution of the data, and is suitable for ordinal data. (GAS) were extracted as measures of affective states to characterize participant's engagement intention. Filtered EEG signals from baseline and different activities were used to compute both an engagement threshold and engagement traces. EEI was the ratio of beta band spectral power (13-22 Hz) to the sum of alpha band spectral power (8-13 Hz) and theta band spectral power (4-8 Hz) [42, 43]. We calculated the EEI at time  $t$  from 40-second sliding window preceding time  $t$ . Bin powers within beta, alpha, and theta bands were summed together to compute the ratio and the ratios from all 14 electrodes were combined to obtain the EEI at time  $t$ . This procedure was repeated every two seconds to generate the

engagement traces. The mean value of the baseline engagement trace was set as the engagement threshold. A summarized EEI was calculated by  $\sum_{i=1}^n (EEI(i) - threshold) / n$  for each activity, where  $n$  is the number of EEI in the related activity.

GAS was computed using preprocessed GSR signal measured from participant’s fingers. Tonic and phasic components were decomposed separately from the raw signal. The signal was first filtered by a 0.5 Hz lowpass filter to remove noise. Then tonic component was acquired by using a 0.05 Hz highpass filter. Phasic component was then calculated by deducting tonic component from the denoised signal. GSR rate, which could be used as arousal state, was calculated by averaging the first derivative of phasic component. For baseline and each activity of the robot experiment, a set of GAS values were calculated using a 40-second sliding window with 38-second overlap. A threshold value was computed by averaging baseline GAS values. Similar to EEI, a summarized GAS was calculated for each activity.

### 3.6.2 One-on-one interaction results

All the participants finished the interaction and completed the surveys and questionnaires (Table 3-1). Cronbach’s alpha coefficients were 0.88 and 0.92, pre- and post-survey respectively. Perceptions became more positive for effort expectancy, attitude, and RAS post-experiment. Wilcoxon signed-rank test results are shown in the table, including the standard score of the Wilcoxon signed ranks, p value, and effect size. It can be seen that attitude subscale and RAS were statistically significantly more positive after HRI at the 0.05 level with medium effect sizes.

Table 3-1. Survey Results for One-on-one Experiment N=11

	Pre <sup>a</sup> M (SD)	Post <sup>a</sup> M (SD)	Z	p	r
Performance Expectancy	25.5 (6.5)	28.6 (7.2)	1.89	0.059	0.40
Effort Expectancy	19.1 (6.6)	15.6 (6.1)	1.73	0.084	0.37
Attitude	50.5 (10.7)	39.7 (13.9)	2.31	0.021	0.49
RAS	100 (21.8)	84.4 (25.0)	2.19	0.028	0.47

<sup>a</sup>Lower values are more positive.

EEI and GAS were computed for 10 participants, because the start/end marker were not recorded properly for the third participant. For each activity and participant, we calculated the corresponding summarized EEI and summarized GAS. The scatter plots (Figure 3-4) illustrates participants’ summarized EEI and summarized GAS with

respect to self-rated activity preferences. In the case of EEI, the dispersion of the data points along each rating level shows that there are individual differences. For example, one participant rated the exercise activity as “extremely interesting (1)” with the summarized EEI of -0.05 whereas another participant provided the same rating with the summarized EEI of 0.18. We further computed the Pearson’s  $r$  to assess the relationship between the summarized EEI and participants’ self-rating data. There was a strong correlation between the two variables ( $r = -0.73$ ,  $N = 27$ ,  $p < 0.001$ ). This strong negative correlation implies that the EEI is high when participants enjoy the activity whereas the EEI is relatively low when participants show less interest to the activity.

Similarly, individual differences were found for GAS. For the music activity and rating level 2, participants’ summarized GAS ranged from -1.05 response peaks/s to 1.66 response peaks/s. No correlation was found between the summarized GAS and self-rated activity preference. The summarized GAS is an important indicator of the intensity of participants’ emotion state. Since the self-rating data indicated the level of likes or dislikes of the activities and were not necessarily associated with changes in arousal states, it is not surprising that no correlation was found. As shown in Figure 3-4, on average participants had a better opinion on the music and exercise activities compared to the other three activities.

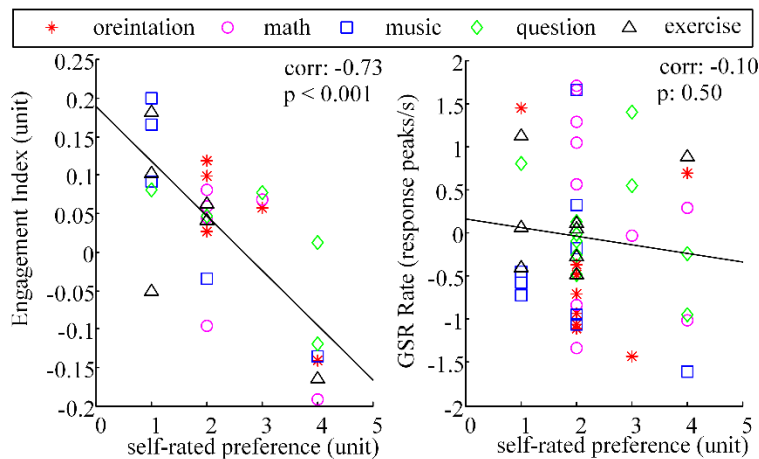


Figure 3-4. Summarized EEI (left) and Summarized GAS (right) as a Function of Self-rated Activity Preferences

### 3.6.3 Triadic interaction results

Survey data were collected for all 14 participants and the results are shown in Table 3-2. Cronbach’s alpha coefficients were 0.93 and 0.92, pre- and post-survey, respectively. All the subscales and RAS indicated more positive perceptions on ROCARE after the experiment. Effort expectancy subscale was statistically significantly more positive after triadic interaction at 0.01 level with a large effect size. Attitude subscale and RAS were statistically significantly more positive after triadic interaction at the 0.05 level with medium effect sizes.



Participants' perceptions on interacting with another person were recorded by a four-item questionnaire. Eleven participants completed this pre- and post-experiment questionnaire. The items were a) I would enjoy doing activities with another person (pre mean score 1.64, post mean score 1.55); b) I would feel comfortable talking to another person (pre mean score 1.36, post mean score 1.18); c) I would help another person when needed (pre mean score 1.18, post mean score 1.27); and d) I would accept help from another person (pre mean score 1.09, post mean score 1.27). While no statistically significant conclusion could be drawn from the questionnaire data, the very small post mean scores show that older adults enjoyed interacting with another person in addition to the robot.

Table 3-2. Survey Results for Triadic Experiment N=14

	<b>Pre <sup>a</sup></b> <b>M (SD)</b>	<b>Post <sup>a</sup></b> <b>M (SD)</b>	<b>Z</b>	<b>p</b>	<b>r</b>
Performance Expectancy	27.9 (9.7)	24.2 (8.4)	1.82	0.069	0.34
Effort Expectancy	18.6 (6.5)	12.9 (7.2)	2.82	0.005	0.53
Attitude	46.4 (15.6)	36.5 (15.0)	2.28	0.023	0.43
RAS	94.8 (30.9)	77.2 (26.4)	2.14	0.033	0.40

We logged EEG data for 6 participants. Their engagement threshold and summarized EEI during the triadic interaction are listed in Table 3-3 together with self-rating data. Each participant had different engagement threshold and different opinions on the “Simon says” activity. Similar to the one-on-one experiment results, the summarized EEI is individualized. Participant 009 rated the “Simon says” activity to be “Somewhat interesting (3)” with an EEI equals 0.27 whereas participant 014’s EEI equals 0.09.

Table 3-3. Summarized EEG Engagement Index for Six Participants

	<b>P009</b>	<b>P010</b>	<b>P011</b>	<b>P012</b>	<b>P013</b>	<b>P014</b>
Threshold	0.95	0.39	0.78	0.65	0.41	0.57
“Simon says” Activity	0.27	0.00	-0.13	-0.20	0.01	0.09
Self-rated Preference	3	2	1	1	1	3

Head pose data were recorded for participants 009 to 014 to analyze whether the participants communicated with each other. The yaw angle results are presented in Figure 3-5. The first row of plots show yaw angles from participants sitting on the right chair (PR), whereas the bottom row plots include data from participants sitting on the left chair (PL). The yaw angle should increase when participants' turn their head to the left and decrease otherwise. Before the triadic interaction, we asked the participants to look at the robot and then look towards the other person. The corresponding head pose yaw angles were horizontal lines in the plots. Pair 007's head pose yaw angles towards each other were not properly recorded. The solid dots in the plots represent instances when the robot provided instructions to elicit HHI. This includes a) acquiring name of PL/PR from PR/PL; b) asking older adults to say hi to each other; c) asking older adults to check each other's gesture if one of them failed; and d) wave goodbye. The number of dots are different for the three pairs because a) and c) might not occur based on real-time human interaction. At the onset of the solid dots, we expect to see that PR's head pose yaw angles increase and PL's head pose yaw angles decrease. From Figure 3-5, this is the case for the majority of the time.

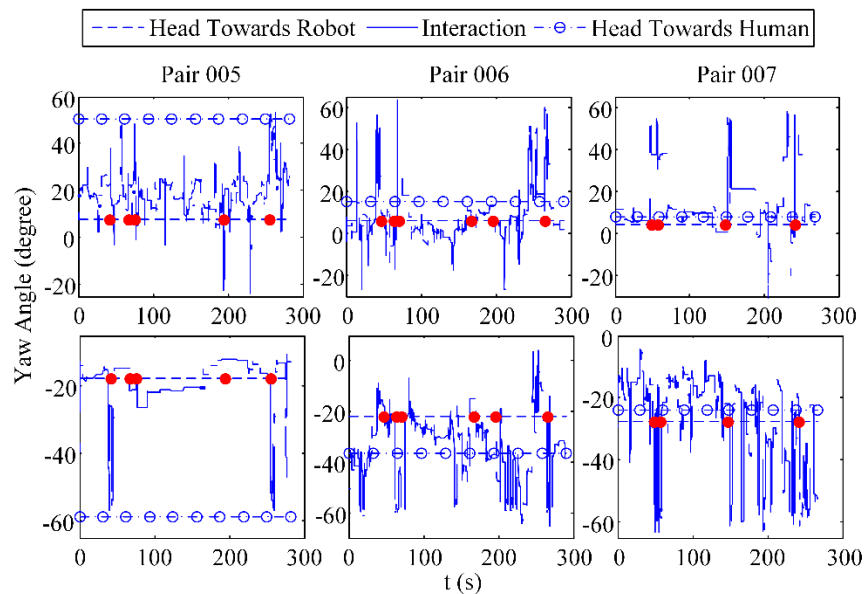


Figure 3-5. Participants' Head Pose Yaw Angles during Triadic Experiment

### 3.7 Discussion

ROCARE is designed to complement and augment care in the existing resource-strained healthcare environment. Several useful interactions between a robot and older adults were developed and tested by small feasibility studies. Overall, one-on-one interaction and triadic interaction systems worked as designed.

No participant dropped out of the studies and the sensors were tolerable. Participants' perceptions after the one-on-one and triadic experiments were significantly more positive on the attitude subscale and RAS. In addition, the survey measurements were sensitive to change from pre- to post-experiment.

EEG and GSR data demonstrated individual differences in baseline features and variation from baseline during HRI. The results also show that the participants had different degree of likes or dislikes of the activities, and therefore it is important for ROCARE to be able to keep track of the preferences of each older adult to maintain engagement. The strong negative correlation between the summarized EEI and participants' self-rating data indicates the potential for objectively measuring participants' engagement intention and harnessing it to realize individualized activity management. Because the correlation was computed using data from all the participants, we cannot be certain that this result applies to each individual in the same way. As for GAS, although no correlation was found between the summarized GAS and self-rated activity preference, it is worthwhile to develop an arousal state-related rating scale and explore the reliability of using GAS as arousal index. In the future, we intend to conduct multi-session experiment and implement the *activity management* submodule using activity preferences learned from electrophysiological signals. This study will provide results on how individual difference affect the EEG and physiological features as well as the effect of activity management.

ROCARE allows for adaptation on two levels of abstraction: a) activity level, which automatically schedules engaging activities; and b) low level, which adapts system behavior based on older adults' real-time interaction, such as gaze and gesture. In this paper, the low level adaptation was implemented. There are several adaptive elements in the system, including EEG and GSR sensors, head pose yaw angle estimation of gaze, and performance related measurements. The EEG and GSR results will serve as the basis for activity level adaptation in the future.

While the current work has demonstrated the potential of a novel HRI architecture with small feasibility studies, it is important to understand the long term effect of such systems in the nursing homes with longitudinal study. Ethical issues and the potential of misuse with robots and older adults have been raised, including decreased human contact, loss of control, loss of privacy, and feelings of objectification [44, 45]. These are serious issues and safeguards need to be considered before their deployment in nursing homes.

### 3.8 Conclusion

Building upon the works of Bohus and Horvitz on multiparty engagement in open-world dialog [36], Louie and associates on multi-user planning and scheduling architecture [24], and state of the art SAR

architectures [20, 28, 30], we proposed the mathematical models of ROCARE, a robotic coach architecture, for augmenting elder care. This architecture is capable of one-on-one and multi-user interactions between a SAR and older adults. By incorporating a database for each individual user and including both implicit and explicit sensing submodules, ROCARE allows individualized activity management and dynamic adaptive robotic behavior for long-term engagement. We have conducted two preliminary feasibility studies: a one-on-one HRI and a triadic HRI. Both systems functioned as desired. Participants' perceptions on the robotic systems were significantly more positive after HRI for the attitude subscale and RAS. Social communication between pairs of participants could be elicited by the robot as seen from both video recordings and head pose data. In addition, there were strong correlation between the summarized EEI and participants' self-rating data ( $r = -0.73$ ,  $p < 0.001$ ), which indicated the potential of using EEG signals for online affective states recognition.

The current work is limited in several ways. First, with the small sample size and the short interaction duration, user perception and compliance results are susceptible to the novelty of the technology. Second, electrophysiology-based affective state recognition was limited to offline analysis. Third, since each participant only took part in one session, no activity preference data were learned and therefore the order of the activities were predefined. Nonetheless, the preliminary studies verified that 1) ROCARE was positively accepted by older adults with and without cognitive impairment; 2) ROCARE can be used for one-on-one and multi-user HRI; and 3) our selection of the EEG feature has strong linear correlation with participants' self-rating on each activity, and can be used for online affective state recognition.

ROCARE is the first to our knowledge that defined multi-user engagement-based mathematical models for robot-mediated interaction for elder care. Future works on building individualized database and activity management need to be carried out with longitudinal studies and a larger sample size. The effectiveness of the architecture to maintain long-term engagement, promote functioning and social communication also needs to be studied.

## REFERENCES

- [1] National Center for Health Statistics, "Older Americans 2012: Key Indicators of Well-being," 2012.
- [2] S. M. Friedman and D. A. Mendelson, "Epidemiology of fragility fractures," *Clinics in geriatric medicine*, vol. 30, pp. 175-181, 2014.
- [3] Alzheimer's Association, "2018 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 14, pp. 367-429, 2018.
- [4] J. N. Vega and P. A. Newhouse, "Mild cognitive impairment: diagnosis, longitudinal course, and emerging treatments," *Current psychiatry reports*, vol. 16, pp. 1-11, 2014.
- [5] V. Freedman, "Disability and care needs of older Americans: An analysis of the 2011 National Health and Aging Trends Study," 2014.
- [6] K. Tabata, "Population aging, the costs of health care for the elderly and growth," *Journal of Macroeconomics*, vol. 27, pp. 472-493, 2005.
- [7] J. Cohen-Mansfield, M. S. Marx, M. Dakheel-Ali, and K. Thein, "The use and utility of specific nonpharmacological interventions for behavioral symptoms in dementia: an exploratory study," *The American Journal of Geriatric Psychiatry*, vol. 23, pp. 160-170, 2015.
- [8] S. C. Forbes, J. P. Little, and D. G. Candow, "Exercise and nutritional interventions for improving aging muscle health," *Endocrine*, vol. 42, pp. 29-38, 2012.
- [9] D. Feil-Seifer and M. J. Mataric, "Defining socially assistive robotics," in *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on*, 2005, pp. 465-468.
- [10] R. Bemelmans, G. J. Gelderblom, P. Jonker, and L. De Witte, "Socially assistive robots in elderly care: A systematic review into effects and effectiveness," *Journal of the American Medical Directors Association*, vol. 13, pp. 114-120. e1, 2012.
- [11] 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*, vol. 74, pp. 14-20, 2013.
- [12] K. Wada, T. Shibata, T. Saito, K. Sakamoto, and K. Tanie, "Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, 2005, pp. 2785-2790.
- [13] H. Gross, C. Schroeter, S. Mueller, M. Volkhardt, E. Einhorn, A. Bley, C. Martin, T. Langner, and M. Merten, "Progress in developing a socially assistive mobile home robot companion for the elderly with mild cognitive impairment," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, 2011, pp. 2430-2437.
- [14] R. Looije, M. A. Neerinx, and F. Cnossen, "Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors," *International Journal of Human-Computer Studies*, vol. 68, pp. 386-397, 2010.
- [15] Y. Matsusaka, H. Fujii, T. Okano, and I. Hara, "Health exercise demonstration robot TAIZO and effects of using voice command in robot-human collaborative demonstration," in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, 2009, pp. 472-477.
- [16] B. Görer, A. A. Salah, and H. L. Akın, "A robotic fitness coach for the elderly," in *International Joint Conference on Ambient Intelligence*, 2013, pp. 124-139.
- [17] M. Simonov and G. Delconte, "Humanoid assessing rehabilitative exercises," *Methods of information in medicine*, vol. 54, pp. 114-121, 2015.
- [18] P. Gadde, H. Kharrazi, H. Patel, and K. F. MacDorman, "Toward monitoring and increasing exercise adherence in older adults by robotic intervention: a proof of concept study," *Journal of Robotics*, 2011.

- [19] K. Yoshino, M. Kouda, and Z. Shanjun, "Correct Motion Advice on Rehabilitation Instruction Robot by Superimposing Instructor CG Model," in *Intelligent Networks and Intelligent Systems (ICINIS), 2012 Fifth International Conference on*, 2012, pp. 333-336.
- [20] J. Fasola and M. J. Matarić, "A Socially Assistive Robot Exercise Coach for the Elderly," *Journal of Human-Robot Interaction*, vol. 2, pp. 3-32, 2013.
- [21] A. Tapus, C. Tapus, and M. Matarić, "Long Term Learning and Online Robot Behavior Adaptation for Individuals with Physical and Cognitive Impairments," in *Field and Service Robotics*, 2010, pp. 389-398.
- [22] D. McColl, W. Y. G. Louie, and G. Nejat, "Brian 2.1: A socially assistive robot for the elderly and cognitively impaired," *Robotics & Automation Magazine, IEEE*, vol. 20, pp. 74-83, 2013.
- [23] W. Y. G. Louie, J. Li, T. Vaquero, and G. Nejat, "A focus group study on the design considerations and impressions of a socially assistive robot for long-term care," in *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*, 2014, pp. 237-242.
- [24] W. Y. G. Louie, T. Vaquero, G. Nejat, and J. C. Beck, "An autonomous assistive robot for planning, scheduling and facilitating multi-user activities," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, 2014, pp. 5292-5298.
- [25] C. D. Kidd and C. Breazeal, "Robots at home: Understanding long-term human-robot interaction," in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, 2008, pp. 3230-3235.
- [26] Z. Zheng, S. Das, E. M. Young, A. Swanson, Z. Warren, and N. Sarkar, "Autonomous robot-mediated imitation learning for children with autism," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, 2014, pp. 2707-2712.
- [27] M. Kanoh, Y. Oida, Y. Nomura, A. Araki, Y. Konagaya, K. Ihara, T. Shimizu, and K. Kimura, "Examination of practicability of communication robot-assisted activity program for elderly people," *Journal of Robotics and Mechatronics*, vol. 23, p. 3, 2011.
- [28] D. Feil-Seifer and M. J. Mataric, "B 3 IA: A control architecture for autonomous robot-assisted behavior intervention for children with Autism Spectrum Disorders," in *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, 2008, pp. 328-333.
- [29] J. Chan and G. Nejat, "Promoting engagement in cognitively stimulating activities using an intelligent socially assistive robot," in *Advanced Intelligent Mechatronics (AIM), 2010 IEEE/ASME International Conference on*, 2010, pp. 533-538.
- [30] J. Chan and G. Nejat, "A learning-based control architecture for an assistive robot providing social engagement during cognitively stimulating activities," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, 2011, pp. 3928-3933.
- [31] H. M. Gross, K. Debes, E. Einhorn, S. Mueller, A. Scheidig, C. Weinrich, A. Bley, and C. Martin, "Mobile Robotic Rehabilitation Assistant for walking and orientation training of Stroke Patients: A report on work in progress," in *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on*, 2014, pp. 1880-1887.
- [32] D. O. Johnson, R. H. Cuijpers, J. F. Juola, E. Torta, M. Simonov, A. Frisiello, M. Bazzani, W. Yan, C. Weber, and S. Wermter, "Socially Assistive Robots: A comprehensive approach to extending independent living," *International Journal of Social Robotics*, vol. 6, pp. 195-211, 2014.
- [33] M. Malfaz, Á. Castro-González, R. Barber, and M. A. Salichs, "A biologically inspired architecture for an autonomous and social robot," *Autonomous Mental Development, IEEE Transactions on*, vol. 3, pp. 232-246, 2011.
- [34] C. Jayawardena and A. Sarrafzadeh, "An alternative approach for developing socially assistive robots," in *Biomedical Robotics and Biomechatronics (2014 5th IEEE RAS & EMBS International Conference on*, 2014, pp. 573-578.
- [35] J. Cohen-Mansfield, M. S. Marx, L. S. Freedman, H. Murad, N. G. Regier, K. Thein, and M. Dakheel-Ali, "The comprehensive process model of engagement," *The American Journal of Geriatric Psychiatry*, vol. 19, pp. 859-870, 2011.

- [36] D. Bohus and E. Horvitz, "Models for multiparty engagement in open-world dialog," in *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 2009, pp. 225-234.
- [37] M. Cummings. (18 November 2015). *Simon Says as a Teaching Tool*. Available: <https://www.youtube.com/watch?v=RewIoHJ9RdM>
- [38] J. Fan, J. W. Wade, D. Bian, A. P. Key, Z. E. Warren, L. C. Mion, and N. Sarkar, "A Step towards EEG-based brain computer interface for autism intervention," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, pp. 3767-3770.
- [39] D. Bian, J. W. Wade, A. Swanson, Z. Warren, and N. Sarkar, "Physiology-based Affect Recognition During Driving in Virtual Environment for Autism Intervention," in *PhyCS*, 2015, pp. 137-145.
- [40] N. Sarkar, "Psychophysiological control architecture for human-robot coordination-concepts and initial experiments," in *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on*, 2002, pp. 3719-3724.
- [41] C. BenMessaoud, H. Kharrazi, and K. F. MacDorman, "Facilitators and barriers to adopting robotic-assisted surgery: contextualizing the unified theory of acceptance and use of technology," *PLoS One*, vol. 6, p. e16395, 2011.
- [42] M. Chaouachi, P. Chalfoun, I. Jraidi, and C. Frasson, "Affect and mental engagement: towards adaptability for intelligent systems," in *Proceedings of the 23rd International FLAIRS Conference*, 2010.
- [43] F. G. Freeman, P. J. Mikulka, L. J. Prinzel, and M. W. Scerbo, "Evaluation of an adaptive automation system using three EEG indices with a visual tracking task," *Biological psychology*, vol. 50, pp. 61-76, 1999.
- [44] N. Sharkey and A. Sharkey, "The eldercare factory," *Gerontology*, vol. 58, pp. 282-288, 2012.
- [45] T. Körtner, "Ethical challenges in the use of social service robots for elderly peopleEthische Herausforderungen zum Einsatz sozial-assistiver Roboter bei älteren Menschen," *Zeitschrift für Gerontologie und Geriatrie*, vol. 49, pp. 303-307, 2016.

## CHAPTER 4

### ROBOT-MEDIATED SOCIAL INTERACTION WITHIN A MULTIMODAL TASK

#### 4.1 Brief Summary

The acceptability and initial results from the triadic HRI experiment described in CHAPTER 3 were promising, which motivated the further development of SAR system for triadic HRI and the development of methods to objectively measure older adults' activity engagement and social interaction during triadic HRI. In this chapter, I describe the design and development of an autonomous robot-mediated interaction system to foster social interaction among older adults within a multimodal task.

This SAR system, SAR-Connect, consisted of three major components: a multimodal task with embedded physical, cognitive, and social stimuli, robot control mechanism to keep older adults engaged in both HRI and HHI, and data analysis algorithms to quantify their social interaction and activity engagement. The multi-modal task is a virtual book sorting activity. We designed a motion-based user interface by means of a Kinect sensor to map older adults' physical movements to manipulative actions in the virtual environment. The task and the mapping rules served as cognitive stimuli. For social engagement, we designed collaborative rules to encourage social communication between older adults. Older adults had to collaborate with each other in order to successfully complete the task. The system monitored real time interaction of older adults, individually and collaboratively, and actively generated robot behaviors to maintain engagement, improve task performance, and encourage collaboration. The system had three operation modes, which were i) interacting with one older adult, ii) interacting with two older adults where they take turns to play the game, and iii) interacting with two older adults where they play simultaneously. The system recorded older adults' interaction data, eye gaze data, vocal sound data, and EEG data continuously in order to evaluate their social interaction and activity engagement.

We recruited older adults to participate in the triadic HRI experiment with two sessions, a one-to-one session and a triadic session. A total number of 26 older adults completed the one-to-one session and 18 of them also completed the triadic session. The experimental results demonstrate that i) the system was able to adapt to older adults' interaction in order to improve their task performance and encourage collaboration; ii) the system was able to automatically measure and evaluate older adults' activity engagement and social interaction, iii) older adults had high activity engagement in the virtual reality-based physically, cognitively, and socially stimulating activity; and iv) older adults had social interaction with each other as induced by the robot-mediated interaction system and their social engagement maintained or slightly increased as they



interacted with the system more. These results indicate that SAR-Connect could be potentially useful to involve multiple older adults to perform multimodal activities with an eye to enhance their functions and foster their social interaction.

There were two research paper published on this work.

- 1) **Fan, Jing**, Linda Beuscher, Paul Newhouse, Lorraine Mion, and Nilanjan Sarkar. “A collaborative virtual game to support activity and social engagement for older adults.” In *International Conference on Universal Access in Human-Computer Interaction*, pp. 192-204. Springer, Cham, 2018.
- 2) **Fan, Jing**, Linda Beuscher, Paul Newhouse, Lorraine Mion, and Nilanjan Sarkar. “SAR-Connect: a socially assistive robotic system to support activity and social engagement of older adults.” *IEEE Transactions on Robotics*, Under Review.

## 4.2 Abstract

Literature suggests that non-pharmacological therapies such as physical, social, and cognitive activities can improve older adults’ overall health and reduce the risk of dementia. Robotic systems can play a key role in providing these activity-oriented therapies in a quantifiable manner to partially mitigate the lack of resources in the healthcare system. While there exist several robotic platforms that can provide some of these activity-oriented therapies, they are primarily limited to one robot and one human interaction and thus do not foster social interaction between multiple humans. In this paper, we present a novel human-robot interaction framework and a realized platform called SAR-Connect to foster robot-mediated social interaction among older adults through carefully designed tasks that also stimulate both physical and cognitive activities. SAR-Connect seamlessly integrates a humanoid robot with a virtual reality-based activity platform and a multimodal data acquisition module including game interaction, electroencephalography, audio, and visual responses of the participants. Results from a user study with older adults showed that this system could 1) involve one or multiple older adults to perform multi-domain activities and provide dynamic guidance, 2) engage them in the robot-mediated task and foster human-human interaction, and 3) quantify their social and activity engagement from multiple sensory modalities. The social and activity engagement results were encouraging indicating the potential of SAR-Connect.

### 4.3 Introduction

In 2014, the number of people aged 65 and over accounted for 15 percent of the population in the US. As the baby boom generation ages, this number will increase dramatically. By 2030, the older population is projected to represent nearly 21 percent of the total population [1]. The majority of the older people has multiple chronic health conditions, which result in increased health care expenditures and limitations in activities of daily living [1, 2]. Among these chronic diseases, dementia is a prevalent syndrome that is characterized by difficulties with memory, language, problem-solving and other cognitive skills related to everyday activities. Alzheimer's disease, the most cause of dementia, is the fifth leading cause of death among older adults. One in ten people age 65 and over has Alzheimer's disease. In addition, approximately 15 to 20 percent of older people have mild cognitive impairment (MCI), a potential precursor to Alzheimer's and other dementias [2].

Physical activity, social engagement, and healthy eating have been shown to improve older adults' physical and psychological well-being and reduce the risks of many health problems [3-5]. Although there is no known cure for Alzheimer's disease, several researchers suggest that engagement in physical, social, and cognitive activities may reduce the risk of Alzheimer's and other dementias [6-8]. Recent reviews of non-pharmacologic therapies indicate that physical activity and cognitive stimulation are beneficial to people with dementia [9, 10]. The costs of health care for older adults especially those with dementia are substantial. Informal caregivers such as family and friends provide the bulk of the care and are under high financial, economic, and emotional burden [1, 2]. Apathy is a prominent behavioral symptom of Alzheimer's dementia. Those with apathy have poor self-motivation, poor initiative, low vitality, non-cooperativeness with care, and diminished goal behaviors that in turn increase the risk for cognitive decline, functional deficits, unsuccessful rehabilitation, social isolation, and caregiver burden and frustration [11, 12]. These lead to the need for technological strategies that can coexist with the current care setting.

To address this need, researchers have developed socially assistive robotic (SAR) systems in order to provide social companionship [13, 14], support independent living [14], facilitate healthy eating [15], and engage older adults in various forms of physical and cognitively stimulating activities [15-20]. Many of these robotic systems are open-loop or remotely operated [21, 22]. Closed-loop SAR systems have the ability to monitor human interaction in real time and adapt system behaviors accordingly and thus are more promising to improve the physical, social, and cognitive health of older adults. Tapus et al. developed a SAR system to help stroke patients and people with cognitive impairment. Two adaptive approaches were presented: on-line adaptation to match robot personality with older adults' preferences, and long-term adaptation to match task difficulty with older adults' task performance [18]. Fasola and Mataric designed a

robotic exercise coach that monitors older adults' performance of chair exercise and actively provide feedback and guidance to encourage task completion [16]. McColl et al. developed a robotic system to engage older adults in meal eating activity and cognitively stimulating activity. The robot adapts its behavior based on the state of the activity and that of the older adult to customize the interaction [15]. While these systems are promising, they were designed and tested to work with a single older adult at one time, and therefore did not address the social aspect of older adults' health and wellbeing by involving multiple users in the activities. In recent years, however, researchers are focusing their attention to one robot to multiple human interaction. Louie et al. developed an autonomous assistive robot that plays bingo game with a group of older adults. However, the goal of the system is to plan and facilitate group activity instead of promoting interpersonal social interaction. Robot's behavior adaptation during game playing was applied only at individual level [19]. Back et al. [20] and Matsusaka et al. [21] developed a SAR system to lead physical activity with multiple older adults. Konah et al. developed a series of robot assisted activities for group interaction [23]. These systems have been shown to be useful however they either operate in an open loop fashion or require a human mediator, and thus are limited in their ability to facilitate social interaction among older adults.

In this work, we seek to further extend SAR's usefulness in fostering social interaction by creating opportunities for two older adults, which can be scaled up for more participants, to interact collaboratively on tasks under the guidance and help from a SAR. This novel platform, called SAR-Connect, uniquely integrates in a closed-loop manner with a virtual reality (VR)-based task environment to present tasks that provide opportunities for physical, cognitive and social interactions as well as a series of data acquisition modalities to quantitatively capture user social interaction and engagement. There are several contributions of this work. *First*, SAR-Connect represents a novel platform for unified multi-domain (i.e., physical, cognitive and social) interaction that can be used with both one-to-one (i.e., one robot and one human) and triadic (i.e., one robot and two humans) modes. Most existing SAR systems are limited to single domain task and do not provide one-to-many interaction opportunities. *Second*, we introduce a VR-based task performance mechanism that is monitored by the SAR in real-time. The VR mechanism allows creation of multitude of controlled tasks in a safe environment considering the unique challenges and disabilities of individual participants and obviate the need for complex sensor systems to capture task performance since the VR software collects these data. *Third*, novel tasks are designed that are embedded with strategies to stimulate physical, cognitive and social interaction. *Fourth*, the SAR is aware of the task performance as well as user engagement and social interaction in real time from the multitude of the data collected by SAR-Connect and thus can generate informed guidance and feedback. Finally, we present experimental data from older adults with and without cognitive impairment to validate SAR-Connect.

The goal of the system is to help older adults remain physically and mentally active, and more importantly, to foster interpersonal social interaction between older adults themselves. The robot is responsible for keeping older adults engaged with the task as well as with each other. *The robotic system provides human-robot interaction (HRI) in both individual and group level, and administers HHI through HRI with a hope to alleviate social isolation and/or loneliness in older adults in the long run.* As the older adults start interacting with each other in the activity-oriented therapies, the role of the robot would gradually fade away. To the best of our knowledge, this is the first instance of a SAR system designed to foster interpersonal social interaction among older adults. The rest of the paper is structured as follows. Section 4.4 describes a few key challenges for our robot-mediated social interaction tasks. Section 4.5 and Section 4.6 present the HRI framework and the design and development of a working SAR system. Section 4.7 and Section 4.8 present the user study and its results. Finally, the results and implications are discussed in Section 4.9.

#### **4.4 Challenges for Robot-mediated Social Interaction**

SAR systems have been developed to administer activity-oriented therapies such as physical exercises and memory games [15, 16]. However, instead of favoring a single modality intervention, the literature on activity-oriented therapies suggests multimodal strategies that are tailored to the individual and highlights the importance of social engagement [5, 24, 25]. Thus, to be most effective, SAR-Connect needs to offer multimodal stimuli, including physical, cognitive, and social components.

In our previous studies, we have explored older adults' perception and acceptance on different forms of physical and cognitive activities as well as simultaneous interaction with their peers [26]. For both one-to-one interaction with the robot and triadic HRI involving two older adults and the robot, older adults' perceptions of the robot were more positive after the session. Social communication between two older adults were observed during triadic HRI. These results indicated that robot-mediated physical and cognitive activities were well tolerated by the older adults and SAR had the potential to involve more than one person and could facilitate interpersonal social interaction. One weakness of our previous robot-mediated activities lies in their ad-hoc nature and the lack of a mechanism to encourage HHI. This brings up the challenge on the design of task structures to involve physical and cognitive stimuli and to enforce or encourage interpersonal social interaction.

The next important question is on the key behaviors of the robot itself. Since the purpose of SAR is to administer activity-oriented therapies and foster social interaction, the robot must have knowledge of older adults' interaction with the task and dynamically guide older adults' to perform the activity and fulfil

task requirements related to the physical, cognitive, and social stimuli. This requires the robot to understand and interpret multi-user HRI in terms of task engagement, performance, and HHI for task completion.

Finally, SAR needs to achieve measurable progress, which includes progress on physical and cognitive functions and that on social interaction. Unlike traditional robotic systems or personal service robots, the progress of older adults cannot be simply extracted from the task specification. Although robot behaviors are tailored to older adults' task performance, performance itself is not a good indicator of older adults' progress due to their vulnerability and the nature of aging. On the other hand, manual analysis of older adults' behaviors by a trained human rater is effort and resource intensive. These require methods to automatically evaluate the progress of older adults during robot-mediated interaction.

#### **4.5 HRI Framework**

We present a HRI framework that will be capable of delivering multimodal intervention strategies through robot-mediated tasks. We first provide a general task structure (Figure 4-1) using hierarchical and modular design, which is flexible enough to accommodate combinations of physical, cognitive, and social stimuli and general enough to account for a large variety of tasks including the ones we have previously developed in [26]. The key components in a task are categorized into physical, cognitive, and social stimuli. Physical stimuli consist of arm movement, hand movement, leg movement, and head movement. Cognitive stimuli involve elements for older adults to perceive, attend, memorize, conceive, and reason. Social stimuli focus on their communication and collaboration during HRI. Subtasks are formed by involving one or multiple key elements. For example, gross motor movement is related to older adults' range of motion (ROM) and is composed of raising arms up, extending arms to the sides, head rotation, and other variation and combination of physical stimuli. Matching, sorting, and question and answer type tasks require key elements from cognitive stimuli. Collaborative rules are mechanisms embedded in the task to encourage older adults to communicate and collaborate with their partners and are related to social stimuli and cognitive stimuli. These subtasks are then combined to form various tasks. For example, a chair exercise is a physical task that is realized by a combination of gross motor movements. Sequence of chair exercise is a physical and cognitive task that combines gross motor movement with matching. Chair exercise with another person will add the social element in the task. It is to be noted that the presented HRI framework is not limited to these task elements alone – it can be expanded to other task elements.

One of the uniqueness of the current HRI framework lies in the seamless integration between a virtual reality (VR) task platform and robot actions. We chose VR for task implementation over a real world task because 1) it is safe to practice in the virtual world, 2) the tasks can be designed to be interesting without

exhausting the persons and without compromising their ROM, 3) instead of requiring older adults' to have the ability and strength (e.g. ROM) to perform a physical task, VR-based interaction can be adjusted to accommodate for differences in older adults' physical ability, and 4) allows comprehensive objective measurements of older adults' interaction. Furthermore, most social robots are not designed to perform physical tasks as their payload is not high. As the functionality and perception capability of social robots become more advanced, VR-based activities could be replaced by a similar physical task without any change of the HRI framework and without sacrificing the advantages provided by the VR as mentioned above.

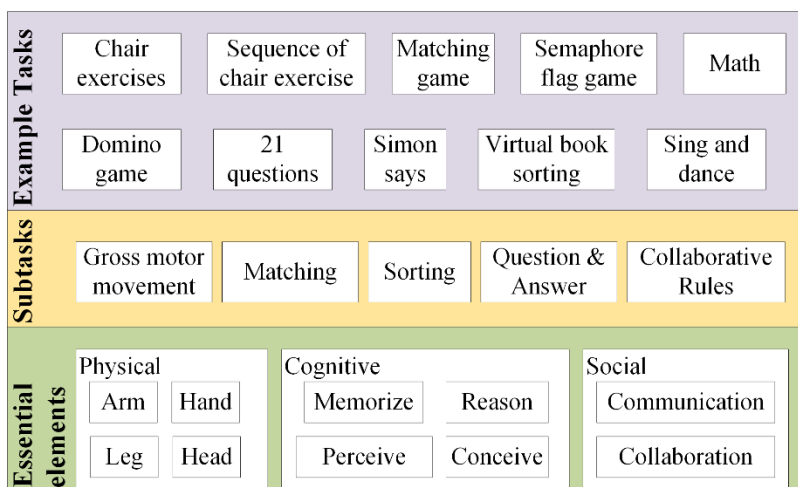


Figure 4-1. Task Structure

We designed a virtual book sorting task to demonstrate the proposed HRI framework that combines gross motor movement, sorting, and collaborative rules to provide physical, cognitive, and social stimuli. The details of the task is discussed in Section 4.6. We chose the humanoid NAO robot ([www.softbankrobotics.com](http://www.softbankrobotics.com)) as the robotic platform to administer the VR-based multimodal task since older adults were interested and engaged in activities led by this robot in our previous studies [26]. However, note that our HRI framework is not limited to work only with the NAO robot. Any robot capable of carrying out complex gestures with an open architecture to integrate with other interactive devices can be used by the framework. In SAR-Connect, the robot is responsible for 1) engaging an older adult with both physical and cognitive exercises; and 2) further helping foster social interaction between two older adults by guiding them to perform collaborative exercises. Specifically, the robot has several roles for the virtual book sorting task. First, it continuously monitors how an older adult is interacting with the system using gestures. Second, the robot observes the state changes in the VR-based task. Third, it guides older adults to achieve task

requirements related to the physical, cognitive, and social stimuli and encourages HHI. And finally, it acts as a user and guidance provider when interacting with a single older adult and takes actions in the VR-based task to perform the activity.

SAR-Connect was composed of a NAO, a Microsoft Kinect for Windows RGB-D sensor, two 14-channel Emotiv electroencephalogram (EEG) headsets (www.emotiv.com), and a VR-based task displayed on a 32 inch computer monitor. One or multiple users sit in front of and facing the Kinect sensor to interact with the system through arm and hand movements. The overall system architecture is shown in Figure 4-2. The Kinect sensor tracks the skeleton positions and hand states of the users and sends them to the *Interaction Manager* module. The *Interaction Manager* module maps the arm and hand movements of the users in real world to hand cursors and grip/release cursor states in the virtual world to allow users to manipulate virtual books. Users' interaction is mediated by NAO via robot speech and gestures. The core element, the *Supervisory Controller* module, communicates with the *Interaction Manager* module, the VR-based multimodal task, and the robot for real-time closed-loop interaction. It gains knowledge on user's interaction with the robot-mediated task by monitoring, updating, and analyzing users' movements, task states, and robot behaviors. It then dynamically guides users' to perform the activity and fulfil task requirements by generating events to trigger robot behaviors as well as audiovisual feedback in the VR-based task.

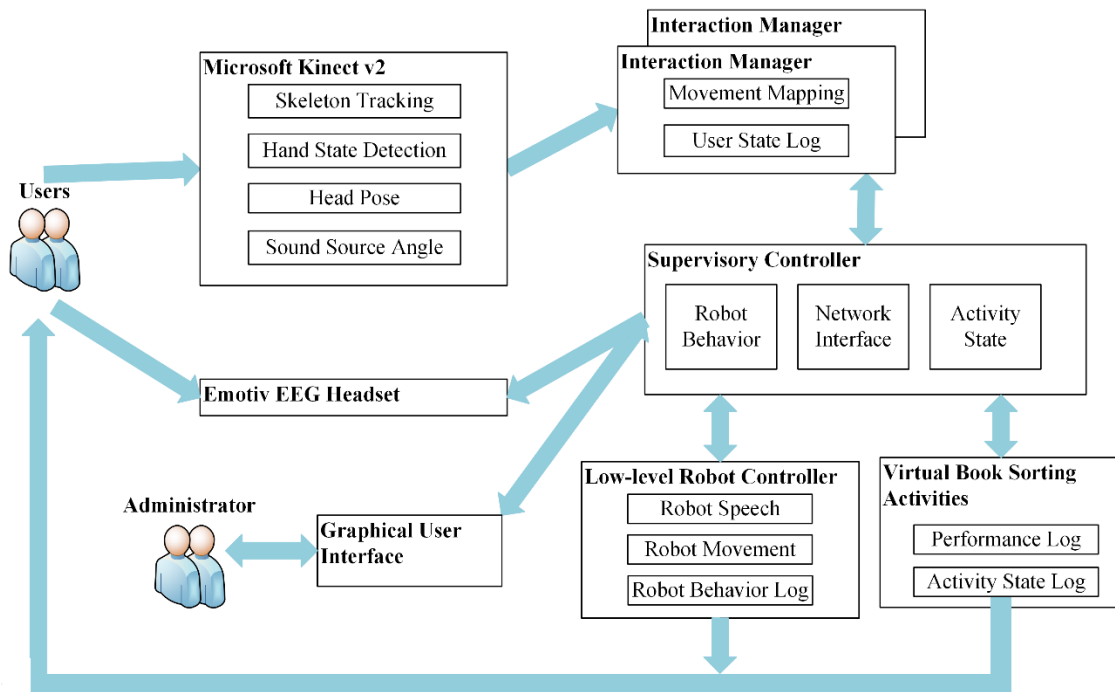


Figure 4-2. System Architecture

In terms of measurable progress, in addition to computing older adults' task performance based on their scores, we added a data acquisition modules in the system to continuously record multimodal sensory data to capture older adults' behaviors during HRI. In addition to the comprehensive interaction data from the VR-based task, the system automatically logs the head pose angles of older adults, the sound source angles detected by the Kinect sensor, robot behaviors, and older adults' EEG signals (Figure 4-2). Since this is the first study to the best of our knowledge of a SAR system to foster interpersonal social interaction, data from multiple sensory modalities were analyzed offline to develop algorithms for the purpose of automatically evaluating social interaction and activity engagement that capture older adults' responses to the physical, cognitive, and social stimuli generated by the robot-mediated task.

## 4.6 System Design and Development

### 4.6.1 VR-based multimodal task

The design details of the VR-based multimodal task was presented in [27]. Here, for the sake of continuity, we briefly describe the virtual book sorting task and its physical and cognitive stimuli. We then discuss for the first time the collaborative rules embedded for encouraging interpersonal social interaction.

The virtual book sorting task was developed using Unity game engine (unity3d.com) and is shown in Figure 4-3. The goal is to sort virtual books into the collection bins based on their colors. Each user will collect books of the color he/she has been assigned to. Efficient collection of some of these books may require help from the other user. By sorting the books as a team, they will increase their game scores. The physical stimuli comes from a motion-based user interface (UI) in the *Interaction Manager* module that naturally maps users' physical movements to manipulative actions in the virtual world. The motion-based UI is realized by means of a Kinect sensor using its skeleton tracking and hand state detection features. It supports grip, move, and release actions through physical movements. To control a hand cursor in the VR-based task by physical movements, we first defined a user's left and right interaction boxes. Figure 4-4 illustrates the front and side views of the interaction boxes in the Kinect coordinate space. Shoulders, hips, and spine joints' positions were used to compute the vertices of the interaction boxes. Only one hand controls one user's hand cursor at a time. The current interaction box is the one that corresponds to the current control hand. In the virtual world, a corresponding 3D interaction area is assigned to each user. A user's cursor position is the projection of his/her hand position from the interaction box in the physical world to the interaction area in the virtual world. Next, we need to allow users' to manipulate books through simple hand gestures. Kinect's hand state detection algorithm returns five possible hand states, which are



closed, lasso, not tracked, open, and unknown. A finite state machine was designed to map these detections to close and open hand states. The detected close and open hand states were then mapped to grip and release cursor states, respectively.



Figure 4-3. Virtual Environment

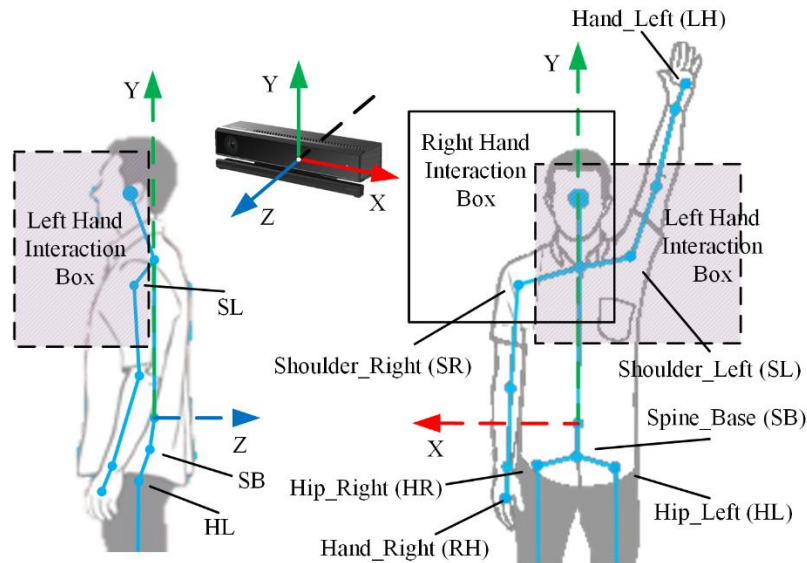


Figure 4-4. Interaction Boxes

The cognitive stimuli are the book sorting task itself, the mapping rules for the physical stimuli, and the collaborative rules for the social stimuli. Two collaborative rules were designed to encourage social communication between older adults. The collaborative rule for red-green book task is shown in Algorithm 1, and the collaborative rule for yellow book task is shown in Algorithm 2. As can be seen, these collaborative rules are not restricted by the number of older adults to allow for future extension of SAR-Connect to work with more than two older adults.

For red-green book task, based on Algorithm 1, the virtual world is divided into two interaction areas, marked by the red and green vertical lines (Figure 4-3). The red interaction area excludes the space to the right of the green vertical line, and the green interaction area excludes the space to the left of the red vertical line. This results in one shared interaction area and two areas that are accessible only by each individual user. The constraints on the interaction areas create the need for collaboration between the users and therefore induce the possibility of social communication. For example, the user controlling the red cursor cannot move the cursor past the green vertical line. Therefore, books that are not reachable by the red cursor need to be moved to the shared interaction area by the user controlling the green cursor. To make the collaborative rule more specific, we defined two collaborative areas, which are the red and green squares on the virtual floor. Users' collaboration is linked with the scoring scheme of the task. Each book has an initial score of 5. The books with numbers on them are called team bonus books, which are positioned far away from the color matched bins. If the user controlling the red cursor moves a green team bonus book closer to the other user by putting the book inside the green collaborative area, this is considered as a collaborative move and the score of the team bonus book increases to 10 accompanied by a rewarding audio feedback. The user is allowed to prevent the other user from scoring by moving any book outside the reach of the other user. Such competitive move is discouraged by resetting the score of the team bonus book back to 5 and playing an error sound.

**Notation:** Subscripts  $s$  and  $w$  denote positions in screen space and world space respectively. Functions  $S2W()$  and  $W2S()$  map positions from screen space to world space and vice versa. Position is represented by Vector3 structure, which contains x, y, z components.

---

**ALGORITHM 1:** Collaborative Rule for Red-Green Book Task

---

**Given:** User's interaction area  $IA_i$ ; user's collaboration area  $CA_i$ .

**Input:** Current book position  $BP_w$ ; user index  $idx$ .

**Requirements for  $IA_i$  and  $CA_i$ :**

$\forall i \neq j, i \in n, j \in n$

{Virtual world is composed of  $IA_i$ ; users share interaction areas}

(1):  $U \leftarrow IA_1 \cup IA_2 \dots \cup IA_n; (IA_i \cap IA_j) \neq \emptyset$

{ $A_i$ : Area that is accessible only by user  $i$ }

(2):  $A_i \leftarrow (U - IA_1 \dots \cup IA_{i-1} \cup IA_{i+1} \dots \cup IA_n) \neq \emptyset; A_i \neq A_j$

(3):  $CA_i \not\subset A_i; CA_i \subset IA_i$

**Update book score given user interactions:**

1: **do forever**

2: **get** start position  $sBP_w$ , end position  $eBP_w$ ,  $idx$

3: **if**  $sBP_w \neq \text{Vector3.zero}$  **and**  $eBP_w \neq \text{Vector3.zero}$

4:  $sBP_s, eBP_s \leftarrow W2S(sBP_w), W2S(eBP_w)$

5: **if** book  $\in$  user  $i$  **and**  $idx \neq i$  **and**  $sBP_s \in U - IA_i$  **and**  $eBP_s \in CA_i$

6: *collaborative move*, book score increases, correct sound

7: **elif** book  $\in$  user  $i$  **and**  $idx \neq i$  **and**  $sBP_s \in IA_i$  **and**  $eBP_s \in U - IA_i$

8: *competitive move*, book score decreases, wrong sound

9: **elif** book  $\in$  user  $i$  **and**  $idx \neq i$  **and**  $sBP_s \in U - IA_i$  **and**  $eBP_s \in IA_i$

10: *unsuccessful collaborative move*

11: **elif** book  $\in$  user  $i$  **and**  $idx = i$  **and**  $sBP_s \in U - IA_i$  **and**  $eBP_s \in IA_i$

12: *reduced chance of collaboration*

---



---

**ALGORITHM 2:** Collaborative Rule for Yellow Book Task

---

**Input:** Users' cursor positions  $CP_s$  and current book position  $BP_w$

**Initialization:**

1:  $prevCP_s[i] \leftarrow \text{Vector3.zero}$  for **for**  $i = 1$  to  $n$

2:  $collaboration \leftarrow [\text{True}, \text{True}, \text{True}]$  {For x, y, z directions}

**Update target position  $TP_w$ :**

3: **do while** users grab the same book

4: **get**  $CP_s, BP_w$

5:  $BP_s, direction, TP_s \leftarrow W2S(BP_w), \text{Vector3.zero}, \text{Vector3.zero}$

6: **for**  $i = 1$  to  $n$  **do** {Calculate  $TP_s$  given each user's  $CP_s[i]$ }

7: **if**  $prevCP_s[i] = \text{Vector3.zero}$

8:  $TP_s \leftarrow n \times BP_s$ ; **break**

9:  $tempDi = CP_s[i] - prevCP_s[i]$

10: **if**  $direction.x = 0$

11:  $direction.x = tempDi.x$

12: **elif**  $collaboration[1]$  **and** ( $direction.x$  matches  $tempDi.x$ )

13:  $collaboration[1] = \text{True}$

14: **else**  $collaboration[1] = \text{False}$

15:  $TP_s.x \leftarrow TP_s.x + CP_s[i].x$  **if**  $collaboration[1]$  **else**  $n \times BP_s.x$

16: **repeat** line 10-15 for  $TP_s$  in y and z directions

17:  $TP_w \leftarrow S2W(TP_s \div n)$ ; Move the book to  $TP_w$

18:  $prevCP_s \leftarrow CP_s$

---

For the yellow book task, based on Algorithm 2, the users collaborate by grabbing the same book and moving the book in the same direction. Otherwise, the book does not move. The moving direction of each user's cursor movement is projected to the X, Y, and Z directions in the virtual world. If the x components of both users are in the same direction, the target position of the yellow book in the X axis is the mean value of the target position of the two hand cursors in the X axis. The target position of the yellow book in the Y and Z axis are computed similarly.

SAR-Connect was developed based on the red-green book task, which we referred to as the main task. The yellow book task was used as the post-test to explore older adults' behaviors when they perform a similar task (book collection) with unknown information. Older adults see yellow books and yellow bins but are not aware of the collaborative rule that they have to move the same book in the same direction together. We are interested in seeing whether older adults would communicate with each other to figure out the unknown piece of the task. If they cannot move any yellow book half way through the interaction, the robot gives them a hint by asking them to try moving together.

#### **4.6.2 Supervisory controller**

The key behaviors of our SAR system were implemented in the *Supervisory Controller* module. During HRI, the system continuously evaluates older adults' activity compliance and collaboration status and generates robot behaviors to engage older adults in the robot-mediated task and social interaction with their peers. There are three interaction modes: one older adult interacts with the system (one-to-one interaction), or two older adults take turns or simultaneously interact with the system (triadic interaction).

Our SAR system is a hybrid system involving both discrete events and continuous dynamics. *Low-level Robot Controller* module and *Virtual Book Sorting Activity* module are responsible for continuous dynamics, including physical behaviors of virtual objects and robot's physical movements. The *Supervisory Controller* module decides the mode transition structure of robot behaviors and activity states. It was modeled by timed automata and hierarchical state machines (HSM), as shown in Figure 4-5. The top level of the hierarchy contains two concurrent states: *robot behavior* and *activity state*. Each of them has a state refinement. These two super-states communicate with each other using shared variables through network interface. Figure 4-5 only illustrates the modes or strategies the system is following. These modes were modeled by HSM, Markov decision processes (MDP), or finite state machines (FSM). The notation for labeling state transitions is *guard / action*. In *Activity State*, as the system receives inputs from *Interaction Manager* and *Low-level Robot Controller* modules, a set of variables, the VE, time bar, and score keep updating. The details of the variables are described shortly in the subsections. *Activity State* is in one of the

three interaction modes. For the mode *triadic interaction – take turns*, the sub-state changes from one user to another when the current interacting user successfully collects a book or makes a collaborative move. The variables associated with the new user are reset with the transition.

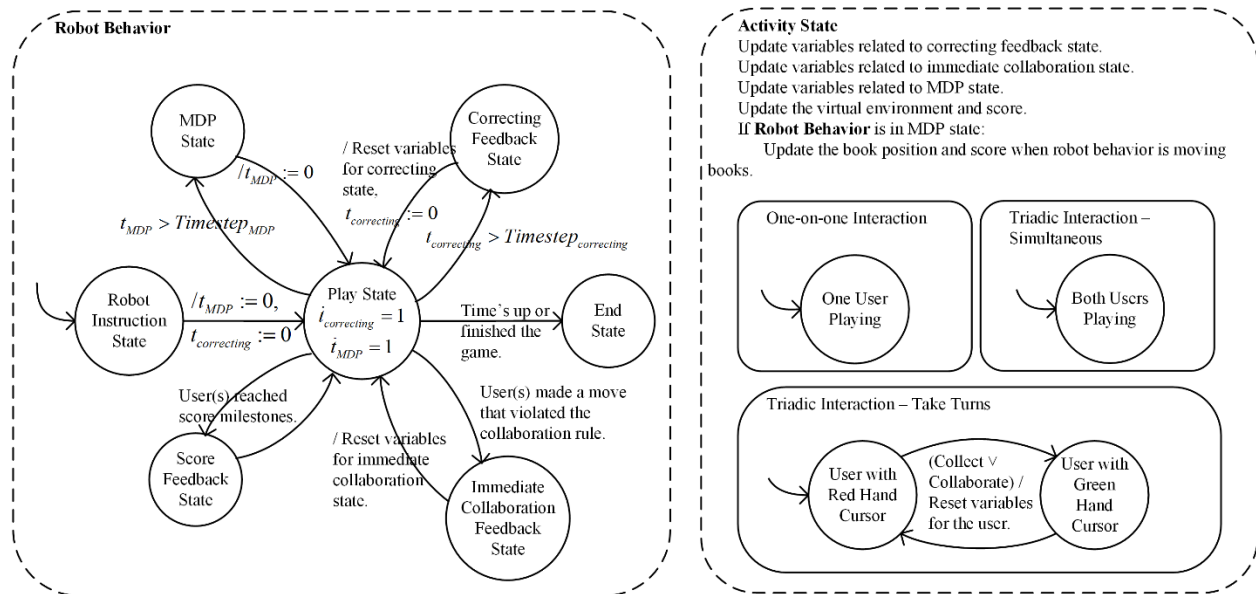


Figure 4-5. Supervisory Controller Module

The initial state for *Robot Behavior* is *Robot Instruction State*, where the robot provides the interaction logic to the user(s). During the interaction, the robot is in one of the five states: *Play State*, *MDP State*, *Correcting Feedback State*, *Immediate Collaboration Feedback State*, and *Score Feedback State*. The robot finishes the interaction in *End State* with a dance if the total score is high. When the robot is in *Play State*, it is in a standing posture with its head rotated towards the VE as if it is monitoring the activity state. *MDP State* and *Feedback States* are the ones that generate robot behaviors automatically based on real time human interactions. State transitions between *Play State* and four other states, *MDP State* and *Feedback States*, are triggered by time variables and discrete events. The robot goes back to *Play State* after *Low-level Robot Controller* completes the designated robot behavior assigned in one of the four states. If no robot behavior is assigned, the robot goes back to *Play State* immediately. While *MDP State* and *Feedback States* are waiting for robot’s continuous dynamics to finish, time variables and discrete events keep updating. We designed the refinement of *Robot Behavior* in this way due to the fact that it takes time for robot to execute any behavior. When the robot is handling one event, another event may occur. This design

ensures that no events are neglected and if one event needs to be handled at a later time, the variables are always up-to-date.

For the purpose of facilitating and engaging older adults to interact with the system and with each other, we developed various robot behaviors ranging from playing as a user to prompting user(s) on how to improve their interaction. These robot behaviors are categorized to *MDP State* or one of the three *Feedback States*. In *MDP State*, the robot plays the role of a second user and assumes that it is playing with a perfect user, who interacts with the system correctly and knows how to obtain a maximal score. In *Feedback States*, the goal of the robot is to make user(s) perform better by prompting user(s) on how to control their hand cursors correctly and how to collaborate with their peers to improve their scores, and by celebrating their score achievement to keep them engaged and motivated. The four states are completely decoupled.

#### 4.6.2.1 MDP state

This state is used only for the one-on-one interaction task, in which the user controls the red hand cursor and the robot acts as a user with the green hand cursor. When a transition from *Play State* to *MDP State* occurs, a MDP model is used to determine the action robot should be taking. The MDP model is a 5-tuple  $(S, A, P, R, \gamma)$ , where:

- $S$  is the finite set of states. These are the current configuration of the VE, and is represented by a 6-tuple  $(c_r, c_g, rb_m, gb_{gm}, gb_r, rb_g)$ .  $c_r$  and  $c_g$  are Boolean values that describe whether the number of collected red books and green books increase or not from time step  $t$  to time step  $t + 1$ , respectively.  $rb_m$  is the number of red books in the red and middle areas, i.e., can be moved by the red hand cursor.  $gb_{gm}$  is the number of green books in the green and middle areas.  $gb_r$  is the number of green books in the red area, and  $rb_g$  is the number of red books in the green area. Given the total number of books  $n$  and assuming equal number of red and green books, the values of  $rb_m$ ,  $gb_{gm}$ ,  $gb_r$ , and  $rb_g$  comply with  $rb_m + rb_g \leq n/2$  and  $gb_{gm} + gb_r \leq n/2$ .
- $A$  is the set of robot behaviors. Four behaviors are defined for *MDP State*, including no action, collect book, offer book, and request book.
- $P: \Pr(s_{t+1} | s_t, a_t)$  is the transition function defining the probability that state  $S$  at time step  $t$  will lead to new state at time step  $t + 1$ , if robot behavior at time step  $t$  is  $a$ . We assume from time step  $t$  to time step  $t + 1$ , robot behavior together with user behavior change state  $S$ . Robot

behaviors are deterministic. However, user behaviors are not. We assume user may take three types of actions: no/failed action, collect book, or offer book. These probabilities model the stochastic user behaviors given robot behavior and were set empirically.

- $R(s_t, a_t, s_{t+1})$  is the immediate reward received after state transition. Positive rewards are given to every book collection and collaborative move made by both user and robot. Negative rewards are associated with impossible robot behaviors under  $s_t$ .
- $\gamma$  is the discount factor that favors immediate rewards over future ones and was set to 0.9.

The time step is at least every  $Timestep_{MDP}$  second. The variables related to *MDP State* are the 6-tuple of state  $S$ .

#### 4.6.2.2 Feedback states

Table 4-1 summarizes the variables related to all *Feedback States* and the resulting robot behaviors. The variables related to the *Correcting Feedback State* were checked in order. If *GripPerc* variable triggered the robot behavior, this indicates *CursorHeight* variable satisfied the requirement and the robot would not evaluate the other two variables. In order to induce social communication between users, in triadic interactions, robot behaviors were designed to target both users instead of one user at a time. For example, if *GripPerc* is low for user A but not user B, robot feedback is “User B, can you help user A with how to grab a book?”

Table 4-1. Variables Related to Feedback States

Feedback States	Variables	Description/Robot Behavior
Correcting Feedback State	CursorHeight	Averaged cursor screen height over time interval $t_{correcting}$ . If this value is below threshold $Th_{ch}$ , NAO reminds users to hold their hands up higher.
	GripPerc	The percentage of time cursor is in grip state over time interval $t_{correcting}$ . If this value is below threshold $Th_{gp}$ , NAO reminds users to close hands and pick up books.
	BookDist	The longest book moving distance during time interval $t_{correcting}$ . If this value is 0, NAO reminds users to move books. Otherwise, if this value is below threshold $Th_{bd}$ , it is usually because the book drops while user is moving it, and NAO suggests users to move books slowly.
	HowToCollab	The number of times users initiate a collaborative move but fail during time interval $t_{correcting}$ . If this value is greater than 0, NAO encourages their attempt and reminds them how to do it correctly.
Immediate Collaboration Feedback State	ShouldCollab	True if users try to collect a team bonus book without collaboration. NAO reminds users to collaborate.
	NoCompete	True if users play competitively by moving a team bonus book outside the reach of the other user. NAO persuades them to stop competing.
Score Feedback State	Score	Cumulative book collection score. NAO celebrates once each time the score is above 30, 60, and 90.

### 4.6.3 Objective measures

We recorded four types of data, which were game interaction, head pose, vocal sound, and EEG, in order to automatically generate objective measures to capture older adults' social and activity engagement. In [26], we defined three engagement variables: engagement action, engagement intention, and engagement state. Engagement actions are older adults' explicit actions related to the task. In SAR-Connect, this corresponds to task-related actions stored in the interaction data and HHI actions in the form of talking to promote task performance, which could be evaluated using vocal sound data. Engagement intention, on the other hand, is implicit states of older adults. In SAR-Connect, this corresponds to where older adults' are paying attention to and their electrophysiological responses. We used older adults' head movements to approximate their gaze and used EEG signals for electrophysiological responses. Together, engagement action and engagement intention determine engagement state based on a timed automaton.

Interaction data indicated older adults' real time interaction with SAR-Connect, including the motion-based control data such as the hand cursors' position and the type and position of a grabbed book, the task

states such as the number of different books in different virtual areas, the performance data such as the total score and the number of collaborative moves, and robot actions. From the interaction data, we defined two metrics to represent activity and social engagement. *Self-effort* was defined as the amount of effort exerted by older adults to move their own books. This included the effort needed to successfully collect a book and the effort needed to move books closer to one's own bin. *Collaboration-effort* was defined as the amount of effort exerted by older adults to help move their peers' books. This included the effort needed to successfully move team bonus books to collaboration area and the effort needed to move peers' book closer to their bins. In this context, the effort was the change of book distance due to older adults' hand and arm movements. These metrics were computed automatically from the motion-based hand control data, the task states, and part of the performance data.

The head pose yaw angles were detected by the Kinect sensor and the data were used to estimate older adults' engagement towards the task and each other in terms of their looking direction during HRI. Before HRI, we recorded about 15s of calibration data where we asked pairs of older adults to look at different locations, including the robot, the computer screen, and at their peers. The calibration data were used to define ranges of head pose yaw angles for head rotation towards the robot, head rotation towards the computer screen, and head rotation towards the other person. Figure 4-6 shows the raw head pose yaw angles for two older adults while they were interacting with the system. The green bands represent the range of head pose yaw angles for looking at the computer screen. The blue bands represent the range of head pose yaw angles for looking at the robot. The red lines indicate the head pose yaw angles for looking at their peers as calculated from the calibration data. In order to include subtle head movements towards another person, instead of setting the red lines as thresholds for looking at another person, we added a margin to the left and right most edges of the system as thresholds for head rotation towards human. These new thresholds are represented as the other edges for head rotation towards human in Figure 4-6. From the head pose yaw angle plots of two older adults, it can be seen that the majority of the yaw angles fall within the ranges of head rotation towards the screen and head rotation towards the robot, and sometimes the yaw angles overshoot or undershoot to reach thresholds for looking at the human. This indicates the ability of the generated ranges and parameters to interpret raw head pose yaw angles as a measure of engagement towards the task and each other. Because the computer screen and the robot were positioned in close proximity and their ranges of head pose yaw angles overlapped with each other, we combined these two ranges into one range of head pose yaw angles for engagement towards the task. These ranges were then used to automatically calculate the amount of time older adults' paying attention to the task and their peers as well as the number of times they looked and/or turned their heads towards each other.



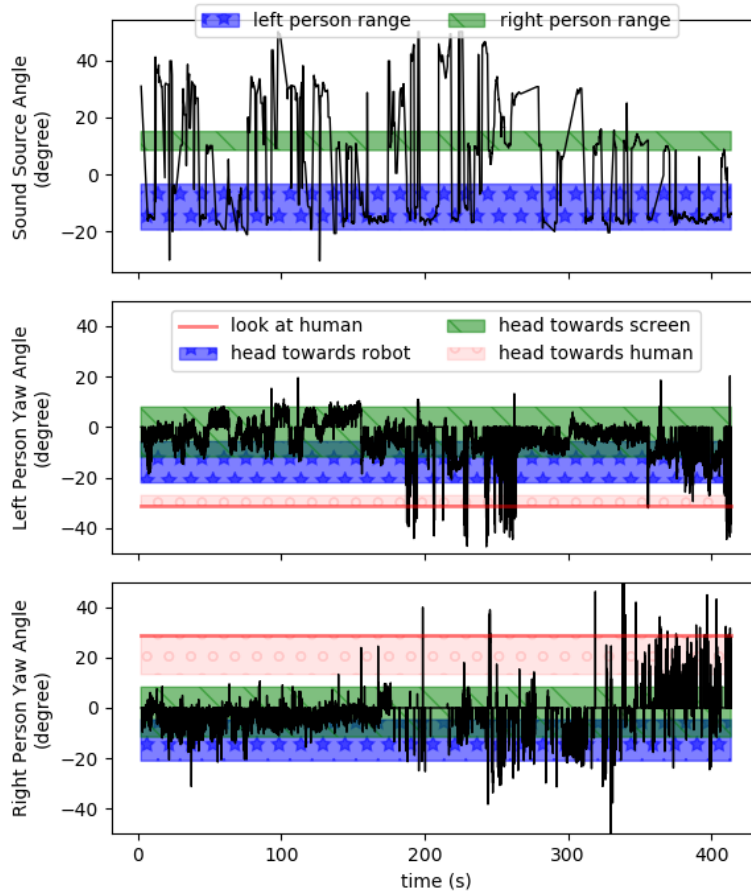


Figure 4-6. Raw Sound Source Angle and Yaw Angle Data for One Pair

In a similar manner, we were able to automatically detect the start and end of vocal sounds made by older adults. The Kinect sensor recorded the sound source angles during HRI, which identified the direction of a sound source. In order to isolate the range of sound source angles for one older adult from the rest of the sound sources such as the other person, the robot, and sounds generated by the virtual task, each older adult was asked to read a sentence during which we recorded the sound source angles and the corresponding confidence levels for the detection before HRI. By aggregating the sound source angle calibration data for all the pairs, we computed the ranges for sound source angles that capture older adults' vocal sounds and the lower bounds for the confidence levels. An example of raw sound source angles recorded during one session of triadic HRI is shown in Figure 4-6. The green band represents the range of sound source angles that detects vocal sounds from the older adult sitting to the right facing the robot. The blue band represents the range of sound source angles that detects vocal sounds from the older adult sitting to the left facing the robot. The sound source angle data that fall outside of these two ranges are detected vocal sounds generated by either the robot, the virtual task, or noise in the environment. These ranges for sound source angles and

parameters for confidence levels of the detection allow us to compute automatically the amount of time older adults were talking and the number of times older adults spoke.

With respect to the EEG signals, we used the EEG engagement index (EEI) [26] to estimate older adults' overall engagement level during HRI. Before HRI, we recorded two minutes of baseline EEG signals during which we asked the participants to sit quietly with eyes open. The EEI was calculated by taking the ratio of beta band spectral power (13-22 Hz) to the sum of alpha band spectral power (8-13 Hz) and theta band spectral power (4-8 Hz). EEIs calculated from baseline EEG signals were averaged to serve as the base engagement level for each older adult. We then computed the summarized EEI during HRI by averaging the change of EEI from baseline for every 40s of EEG epoch. Details of the algorithms to process EEG signals and to compute EEI and summarized EEI are described in our previous papers [26, 28].

## **4.7 User Study**

### **4.7.1 Experimental design**

SAR-Connect, the experimental room setup, and the experimental procedure are shown in Figure 4-7. Participants sat in the two chairs in front of and facing the system. NAO was positioned by the side of the computer monitor. The Kinect was placed on the edge of the table in front of the monitor. Participants sat approximately 2 meters away from the monitor and at a 30 degree angle towards each other. When a single participant interacted with the robot, one chair was positioned directly in front of the table. The experimental procedure consisted of five components: a practice session, which is then followed by three main tasks (one-to-one HRI, triadic HRI – take turns, and triadic HRI – simultaneous), and finally a post-test. Each participant first interacted independently with the system and then pairs of participants played with each other under the guidance of NAO. During practice, the robot taught participants how to interact with the system by arm movement and hand manipulation. Participants then performed the main task alone with the robot as the second player. After two older adults completed the one-to-one HRI, they were paired to perform the main task together. They first took turns to interact with the system and then played again simultaneously. Lastly, they completed the post-test to finish the whole experiment.

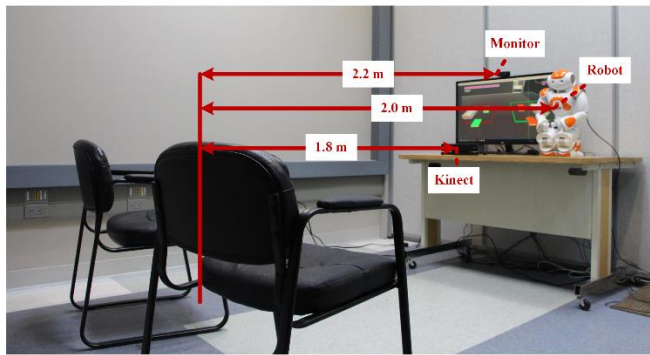


Figure 4-7. Experimental Setup and Procedure

For the main tasks, there were 8 red books and 8 green books, and 10 out of the total 16 books were team bonus books. If none of the participants collaborated, the maximum score they could obtain was 80. If they collaborated for every team bonus book, the maximum score increased to 130. The robot encouraged them to achieve high score in *Robot Instruction State* by telling them it would dance to celebrate if they had achieved more than 100 points. The duration of the interaction excluding *Robot Instruction State* and *End State* was limited to six minutes. The thresholds for *Feedback States* were set to be  $Th_{ch} = 100$ ,  $Th_{gp} = 0.1$ , and  $Th_{bd} = 2$ .  $Timestep_{MDP}$  was 6s if participants finished their part and were waiting for the robot to complete the task. Otherwise,  $Timestep_{MDP}$  was 12s.  $Timestep_{correcting}$  was 15s for triadic interactions – take turns and 20s for the other two main tasks. For the post-test, the interaction duration was set to be three minutes. All these parameters were chosen based on limitation of NAO (runs about 10~15min before motors become hot) and by pilot testing with older adults and volunteers.

#### 4.7.2 Participants

The study was approved by the Vanderbilt University Institutional Review Board. We separated the five tasks into two sessions: a one-to-one session and a triadic session. Participants came to the laboratory for practice and one-to-one interaction first. If they completed the tasks and were willing to come back for another session, we paired them with another participant who had finished the one-to-one session. The triadic session consisted of the two triadic interactions and the post-test. A total number of 26 older adults

took part in the study (17 females, 9 males). The age of the participants ranged from 70 to 90 years old (Mean = 76.7, SD = 5.6). The Montreal cognitive assessment (MoCA© Version 7.1) was used as a screening tool for MCI and early Alzheimer's dementia [29]. Participants' MoCA score ranged from 19 to 27 (Mean = 24.2, SD = 2.2). Nine participants had normal cognition, 12 had MCI, and 5 had Alzheimer's dementia. Out of the 26 older adults, 18 were paired for triadic interaction (7 had normal cognition, 8 had MCI, and 3 had Alzheimer's dementia). For one of the pairs, both older adults had severe hearing issues and were not able to understand the robot. It was even difficult for them to understand the administrator. We thus removed this pair's data. The rest of the older adults who dropped out were mostly due to scheduling issue.

## 4.8 Results

### 4.8.1 System performance results

The system worked as designed. The VR-based tasks were displayed and updated correctly. The motion-based UI was stable and older adults could easily move their hands to control their hand cursors in both horizontal and vertical directions. There were times participants struggled to move books in the third direction, which corresponds to depth in the virtual environment. However, once they learned this type of motion, they were able to perform it without help from the administrator. During one-to-one HRI, triadic HRI, and post-test, there was no interruption by the administrator unless the older adults were unable to interact with the system and became very frustrated. This rarely happened during the experiment.

In one-to-one HRI, the robot was able to play and make progress towards task completion for all the participants. In both one-to-one HRI and triadic HRI, the robot prompted older adults on their task performance and encouraged older adults to collaborate with each other following the *Supervisory Controller* module. All robot behaviors generated by the *Supervisory Controller* module were executed successfully. In addition to activity instructions and celebration feedback generated by the robot, a total number 513 robot behaviors were generated. In one-to-one HRI, 204 robot behaviors were collecting books ( $M = 8.50$ ,  $SD = 2.48$ ), 109 robot behaviors were offering books ( $M = 4.54$ ,  $SD = 0.78$ ), 8 robot behaviors were requesting books ( $M = 0.33$ ,  $SD = 0.70$ ), 23 robot behaviors were increasing task performance ( $M = 0.96$ ,  $SD = 1.23$ ), and 50 robot behaviors were increasing collaboration ( $M = 2.08$ ,  $SD = 1.47$ ). In take turns HRI, 41 robot behaviors were increasing task performance ( $M = 5.13$ ,  $SD = 3.48$ ) and 28 robot behaviors were increasing collaboration ( $M = 3.50$ ,  $SD = 2.83$ ). In simultaneous HRI, 19 robot behaviors were increasing task performance ( $M = 2.38$ ,  $SD = 1.92$ ) and 31 robot behaviors were increasing collaboration ( $M = 3.88$ ,  $SD = 1.55$ ). Since there were 8 books in total and

5 of them were team bonus books, an average of 8.5 collecting book behavior and 4.54 offering book behavior indicate that the robot was trying to collect all the books it could collect and trying to offer help as much as possible. Thus, the robot was able to play the game with older adults. In terms of task performance and collaboration feedback, the number of robot prompts were in line with older adults' interaction data, which we present in the following section. When older adults performed worse, the robot provided more prompts. The system logged all the generated robot behaviors, activity states, participants' interaction data, head pose data, vocal sound data, and EEG signals correctly.

#### **4.8.2 Objective measures of interaction**

From the interaction data, we computed the self-effort and collaboration-effort metrics. The collaboration-effort is related to social engagement between the two older adults whereas the combination of two efforts is related to their activity engagement. In one-to-one HRI, the collaboration-effort is computed in the same way to measure amount of collaboration between the robot and the human. Figure 4-8 and Table 4-2 show participants' effort analysis results during one-to-one and triadic HRI. We used the Wilcoxon signed rank test to compare effort analysis results. One-to-one HRI had the highest collaboration-effort and total-effort. This result is expected since the robot was designed to perform as a collaborative player as well as to prompt older adults on their task performance and encourage them to collaborate. In addition, during triadic HRI, participants interacted with both of their peers and the system and helped each other rather than only focused on their own task. As a result, divided attention is likely to play a role in the decrease of efforts. For triadic HRI, participants' collaboration-effort and total-effort increased from take turns HRI to simultaneous HRI. As compared to take turns HRI, both the collaboration-effort ( $Z = 2.07, r = 0.37$ ) and the total-effort ( $Z = 2.28, r = 0.40$ ) in simultaneous HRI were statistically significantly higher at the 0.05 level with a medium effect size. The increase of total-effort was partially due to a statistically significant increase of self-effort in simultaneous HRI as compared to take turns HRI ( $Z = 2.28, r = 0.40$ ). However, the increase in collaboration-effort and total-effort need to be interpreted cautiously. Since the two triadic HRI tasks were not exactly the same, although these increased efforts are positive and indicate the potential usefulness of the system, we were not able to conclude from these two metrics that older adults' activity and social engagement increased as they interacted with the system more. We also calculated the ratio of the self-effort to the collaboration-effort. The result of the ratios is listed in Table 4-2. This result shows that older adults had more self-effort compared to collaboration-effort for both one-to-one HRI and triadic HRI. The differences among the ratios for the three types of HRI were not statistically significant. Therefore, when older adults' collaboration-effort increased or decreased from one

task to another, their self-effort followed the change. This indicates that older adults' maintained their collaboration-effort throughout the entire HRI.

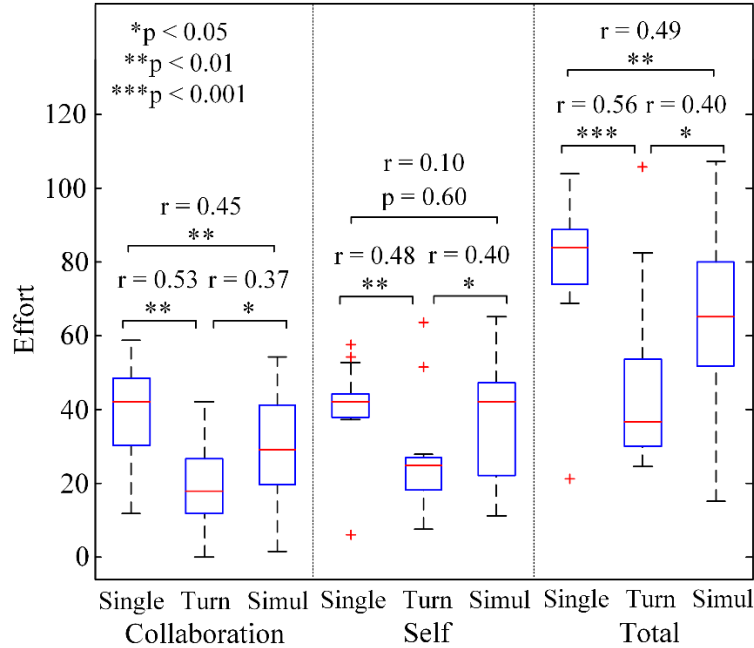


Figure 4-8. Effort Analysis Results

Table 4-2. Effort Analysis Results

Performance Metric	One-to-One		Take Turns		Simultaneous	
	Mean	SD	Mean	SD	Mean	SD
Self-effort (SE)	41.34	11.22	25.52	14.24	37.32	16.18
Collaboration-effort (CE)	39.08	13.53	18.93	11.16	29.03	15.77
Ratio (SE/CE)	1.26	1.03	1.68	1.48	2.28	2.56
Total effort	80.42	18.87	44.45	22.35	66.36	24.29

In terms of the post-test task, 5 out of 8 pairs of older adults were able to figure out the unknown collaborative rule through social interaction without help from the robot. For successful collaboration, we computed the number of times older adults moved a yellow book together, the amount of effort older adults exerted to move a yellow book together, and the amount of time they spent to move a yellow book together. For task effort that is unsuccessful, we computed the number of times older adults tried to move a yellow book and the amount of time they spent to move a yellow book. The total duration takes into account both successful collaboration as well as unsuccessful task effort. The results are listed in Table 4-3. It can be

seen from the table that on average, older adults successfully moved yellow books 6 times and spent 75.7s actively engaging in the 3-minute task. Although the group that did not receive a hint from the robot had more successful collaboration and more unsuccessful attempts, the total durations for the two groups are comparable. These results indicate participants' engagement in an unseen task and their willingness to interact with each other and explore the collaborative rule.

Table 4-3. Post-test Results

Performance Metric	All		Without Hint		With Hint		
	Mean	Range	Mean	Range	Mean	Range	
Success	Count	6.4	[2, 12]	7.8	[2, 12]	4.0	[3, 6]
	Effort	6.8	[1.8, 14.1]	8.8	[2.1, 14.1]	3.6	[1.8, 6.0]
	Duration (s)	38.9	[7.0, 97.0]	44.0	[7.0, 97.0]	30.2	[16.5, 42.8]
Fail	Count	16.9	[5, 32]	18.7	[6, 32]	13.8	[5, 21]
	Duration (s)	36.8	[15.7, 96.7]	30.7	[15.7, 83.5]	46.9	[20.6, 96.7]
Total Duration (s)	75.7	[27.7, 118.5]	74.8	[27.7, 115.9]	77.2	[58.9, 118.5]	

From the head pose data, we computed the amount of time older adults were paying attention to the computer screen or the robot as activity engagement. For social engagement, we computed the amount of time and the number of times older adults' looked towards their peers. The results are listed in Table 4-4. In general, older adults' spent the majority of the time (80.7% in take turns HRI, 75.9% in simultaneous HRI, and 86.2% in post-test) focusing on the task and the system. These data indicate their overall engagement on triadic HRI. They also had social engagement in terms of looking towards their peers. In take turns HRI, the number of times they looked towards their peers ranged from 0 to 25 (median: 2.5). In simultaneous HRI, the number of times they looked towards their peers ranged from 0 to 19 (median: 2). In post-test task, the number of times they looked towards their peers ranged from 0 to 9 (median: 1). On average, older adults looked at their peers 0.7 times per minute in triadic HRI and 0.6 times per minute in post-test. Compared to take turns HRI, older adults looked towards their peers more during simultaneous HRI. The percentage of looking time duration increased from 4.1% to 5.6%. As a result, their activity engagement decreased slightly. In terms of the post-test, older adults spent less time on looking behaviors but they spent more time focusing on the system. None of the changes in activity engagement or social engagement are statistically significant based on the Wilcoxon signed rank test.

Table 4-4. Head Pose Analysis Results

Head Pose Metric		Take Turns		Simultaneous		Post-test	
		Mean	SD	Mean	SD	Mean	SD
Activity Engagement	Duration (s)	309.3	43.0	253.6	70.5	155.2	26.5
	Percentage	80.7%	11.3%	75.9%	14.4%	86.2%	14.7%
Social Engagement	Duration (s)	16.0	26.1	19.3	31.5	3.9	5.8
	Percentage	4.1%	6.8%	5.6%	8.6%	2.2%	3.2%
	Count	4.7	6.5	3.8	4.7	1.9	2.3
	Count per minute	0.7	1.0	0.7	0.8	0.6	0.8

From the vocal sound data, we computed the total amount of time and the number of times older adults were speaking. The results are listed in Table 4-5. For individual older adults, they spent about 12% of the time talking to each other during triadic HRI. In post-test, the amount of talking increased significantly, nearly doubled (21.5%) as compared to triadic HRI. This increase is statistically significant (Wilcoxon signed rank test) for both take turns HRI ( $Z = 2.90, r = 0.51$ ) and simultaneous HRI ( $Z = 2.84, r = 0.50$ ) at the 0.01 level with a medium to large effect size. This is expected since the only way for older adults to figure out how to move a yellow book is through trial and error and communication. For the older adult pairs, the standard deviation of the amount of talking decreased significantly as compared to individual results. This is due to the fact that in most cases older adults' talking were not balanced. One older adult would talk more while the other talked less. It can be seen that older adults talked slightly more in simultaneous HRI than take turns HRI. By the time they performed the post-test, their talking between each other were more balanced. Collectively, these results are positive and indicate that older adults were engaged in social interaction during the entire session of HRI and the slight improvements indicate that the system might be potentially useful.

The summarized EEI was used to estimate older adults' overall engagement level during HRI. We list the results in Table 4-6. As can be seen from the table, the summarized EEIs were comparable for different type of HRI given the large standard deviation as compared to the mean values. Out of 16 older adults, 9 older adults' engagement level increased in triadic HRI than one-to-one HRI. We also calculated the mean and standard deviations of the summarized EEI for the increased group and the decreased group. The results indicate that despite whether older adults preferred triadic HRI or one-to-one HRI, their engagement level as estimated by the summarized EEI increased as they continued interacting with the system during triadic HRI.



Table 4-5. Vocal Sound Analysis Results

Vocal Sound Metric		Take Turns		Simultaneous		Post-test	
		Mean	SD	Mean	SD	Mean	SD
Individual	Duration (s)	45.9	42.3	43.6	44.1	38.6	27.2
	Percentage	11.9%	10.9%	12.6%	11.8%	21.5%	15.1%
	Count	34.7	28.5	31.9	29.0	26.3	14.5%
	Count per minute	5.4	4.4	5.6	4.7	8.8	4.8
Pair	Duration (s)	91.8	51.1	87.3	66.1	77.3	28.1
	Percentage	23.7%	13.2%	25.3%	16.9%	42.9%	15.6%
	Count	69.4	41.7	63.9	43.2	52.5	18.9
	Count per minute	10.8	6.5	11.2	6.7	17.5	6.3

Table 4-6. EEG Analysis Results

Summarized EEI	One-to-One		Take Turns		Simultaneous		Post-test	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Increased Group	-0.086	0.106	0.004	0.100	0.005	0.116	0.024	0.114
Decreased Group	0.006	0.151	-0.136	0.144	-0.134	0.173	-0.119	0.167
Whole Group	-0.049	0.129	-0.052	0.135	-0.050	0.153	-0.033	0.115

## 4.9 Discussion and Conclusion

We present a novel SAR system that aims to augment the care of older adults with or without cognitive impairment with an eye to foster interpersonal social interaction. The ultimate goal of the system is to improve the physical, social, and cognitive aspects of older adults' health and wellbeing. After identifying the key behaviors and properties of such SAR systems, we designed a HRI framework specifically for robot-mediated social interaction among older adults. We then implemented an instance of such SAR systems to provide multimodal stimuli and foster interpersonal social interaction. A virtual book sorting task is carefully designed with embedded physical, social, and cognitive stimuli. The physical stimuli comes from a naturalistic mapping from older adults' physical movement to control inputs in the robot-mediated task. The book sorting task together with the embedded collaborative rules form the cognitive and social stimuli. The system monitors real time interaction of one or multiple older adults and actively generates robot behaviors to maintain engagement, improve task performance, and encourage collaboration. Quantitative data are recorded from multiple sensory modalities to serve as measures of older adults' activity engagement and social interaction during HRI. System testing results from a user study conducted with the target population are promising. System performance results indicate the usability and older adults'

acceptance of the motion-based UI, the task, and the integrated SAR system. From the robot behavior data, it can be seen that the system was able to perform the task with older adults, and adapt to their interaction in order to improve their task performance and encourage collaboration.

We developed a suite of data analysis methods to quantify HRI and HHI from multiple sensory modalities, including task interaction data, head pose, vocal sound, and EEG. The results from these data demonstrate the ability of the system to automatically measure older adults' progress in terms of social and activity engagement during HRI. For triadic HRI, participants had high activity engagement based on the head pose analysis results, and their engagement level increased as they interacted more with the system based on the EEG analysis results. In addition, participants' social engagement maintained or slightly increased as they interacted with the system more based on the head pose and the vocal sound data analysis results. From the interaction data analysis results, participants' collaboration-effort and total task effort increased as they interacted with the system more. Due to the fact that older adults only took part in a single session, no conclusion could be drawn from these results on whether the system is able to enhance older adults' functions and foster their social interaction. However, the activity engagement and social interaction results from different sensory modalities have no contradictions and are positive, which support further investigation on the efficacy of such a system. Participants' post-test performance, activity and social engagement further show promising results that their collaboration behavior transferred to an unseen task with unknown collaborative rule, and their social interaction and activity engagement from different sensory modalities either maintained or increased.

The presented SAR system can engage one or multiple older adults in a closed-loop fashion. In the case of one robot to multiple older adults, the system adapts to both individual and group interaction. More importantly, in addition to HRI, the system is specifically designed to target the interaction between two older adults, i.e., HHI. HHI is the key for such SAR systems to alleviate social isolation and/or loneliness in older adults and in turn increase their motivation and engagement in activity-oriented therapies or rehabilitations, either with a robotic system or a healthcare provider. In this work, the robot monitors and guides HHI mainly by evaluating human collaboration. This system demonstrated the possibility of inducing HHI through HRI. More sophisticated system adaptation based on HHI can be embedded to the presented HRI framework.

We believe it is more beneficial for an older adult to interact with a human being in activity-oriented therapies than with a machine. Such robot-mediated systems with the capability to induce and further shape HHI in the long run would gradually reduce its role in the task and provide only necessary guidance as the users start interacting and engaging in the task and with each other. A HRI framework that not only engages

older adults in physical and cognitive exercises but also fosters interpersonal social interaction between older adults has great potential to improve their physical and psychological well-being.

The current system is limited in that only interaction data and collaboration in task were evaluated to adapt robot behaviors. In the future, we intend to include the activity engagement and social interaction measures as a way to evaluate real-time interpersonal social interaction and task engagement, and extend the adaptive behaviors of the robot to shape the social interaction among older adults. We will also design tasks with different difficulty levels to accommodate older adults with different cognition level.

## REFERENCES

- [1] Federal Interagency Forum on Aging-Related Statistics, "Older Americans 2016: Key Indicators of Well-Being," *Federal Interagency Forum on Aging-Related Statistics*, 2016.
- [2] Alzheimer's Association, "2017 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 13, pp. 325-373, 2017.
- [3] J. Reedy, S. M. Krebs-Smith, P. E. Miller, A. D. Liese, L. L. Kahle, Y. Park, and A. F. Subar, "Higher diet quality is associated with decreased risk of all-cause, cardiovascular disease, and cancer mortality among older adults," *The Journal of nutrition*, vol. 144, pp. 881-889, 2014.
- [4] W. J. Chodzko-Zajko, D. N. Proctor, M. A. F. Singh, C. T. Minson, C. R. Nigg, G. J. Salem, and J. S. Skinner, "Exercise and physical activity for older adults," *Medicine & science in sports & exercise*, vol. 41, pp. 1510-1530, 2009.
- [5] J. Cohen-Mansfield, M. S. Marx, M. Dakheel-Ali, and K. Thein, "The use and utility of specific nonpharmacological interventions for behavioral symptoms in dementia: an exploratory study," *The American Journal of Geriatric Psychiatry*, vol. 23, pp. 160-170, 2015.
- [6] A. Karp, S. Paillard-Borg, H.-X. Wang, M. Silverstein, B. Winblad, and L. Fratiglioni, "Mental, physical and social components in leisure activities equally contribute to decrease dementia risk," *Dementia and geriatric cognitive disorders*, vol. 21, pp. 65-73, 2006.
- [7] C. Hall, R. Lipton, M. Sliwinski, M. Katz, C. Derby, and J. Verghese, "Cognitive activities delay onset of memory decline in persons who develop dementia," *Neurology*, vol. 73, pp. 356-361, 2009.
- [8] H.-X. Wang, W. Xu, and J.-J. Pei, "Leisure activities, cognition and dementia," *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, vol. 1822, pp. 482-491, 2012.
- [9] N. Farina, J. Rusted, and N. Tabet, "The effect of exercise interventions on cognitive outcome in Alzheimer's disease: a systematic review," *International Psychogeriatrics*, vol. 26, pp. 9-18, 2014.
- [10] E. Aguirre, R. T. Woods, A. Spector, and M. Orrell, "Cognitive stimulation for dementia: a systematic review of the evidence of effectiveness from randomised controlled trials," *Ageing research reviews*, vol. 12, pp. 253-262, 2013.
- [11] S. Ishii, N. Weintraub, and J. R. Mervis, "Apathy: a common psychiatric syndrome in the elderly," *Journal of the American Medical Directors Association*, vol. 10, pp. 381-393, 2009.
- [12] E. H. Hölttä, M.-L. Laakkonen, J. V. Laurila, T. E. Strandberg, R. S. Tilvis, and K. H. Pitkälä, "Apathy: prevalence, associated factors, and prognostic value among frail, older inpatients," *Journal of the American Medical Directors Association*, vol. 13, pp. 541-545, 2012.
- [13] R. Yu, E. Hui, J. Lee, D. Poon, A. Ng, K. Sit, K. Ip, F. Yeung, M. Wong, and T. Shibata, "Use of a Therapeutic, Socially Assistive Pet Robot (PARO) in Improving Mood and Stimulating Social Interaction and Communication for People With Dementia: Study Protocol for a Randomized Controlled Trial," *JMIR research protocols*, vol. 4, 2015.
- [14] H. Gross, C. Schroeter, S. Mueller, M. Volkhardt, E. Einhorn, A. Bley, C. Martin, T. Langner, and M. Merten, "Progress in developing a socially assistive mobile home robot companion for the elderly with mild cognitive impairment," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, 2011, pp. 2430-2437.
- [15] D. McColl, W. Y. G. Louie, and G. Nejat, "Brian 2.1: A socially assistive robot for the elderly and cognitively impaired," *Robotics & Automation Magazine, IEEE*, vol. 20, pp. 74-83, 2013.
- [16] J. Fasola and M. J. Matarić, "A Socially Assistive Robot Exercise Coach for the Elderly," *Journal of Human-Robot Interaction*, vol. 2, pp. 3-32, 2013.
- [17] B. Görer, A. A. Salah, and H. L. Akın, "A robotic fitness coach for the elderly," in *International Joint Conference on Ambient Intelligence*, 2013, pp. 124-139.
- [18] A. Tapus, C. Tapus, and M. Matarić, "Long Term Learning and Online Robot Behavior Adaptation for Individuals with Physical and Cognitive Impairments," in *Field and Service Robotics*, 2010, pp. 389-398.

- [19] W. Y. G. Louie, T. Vaquero, G. Nejat, and J. C. Beck, "An autonomous assistive robot for planning, scheduling and facilitating multi-user activities," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, 2014, pp. 5292-5298.
- [20] I. Bäck, K. Makela, and J. Kallio, "Robot-Guided Exercise Program for the Rehabilitation of Older Nursing Home Residents," *Annals of Long-Term Care: Clinical Care and Aging*, vol. 21, pp. 38-41, 2013.
- [21] Y. Matsusaka, H. Fujii, T. Okano, and I. Hara, "Health exercise demonstration robot TAIZO and effects of using voice command in robot-human collaborative demonstration," in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, 2009, pp. 472-477.
- [22] R. Looije, M. A. Neerincx, and F. Cnossen, "Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors," *International Journal of Human-Computer Studies*, vol. 68, pp. 386-397, 2010.
- [23] M. Kanoh, Y. Oida, Y. Nomura, A. Araki, Y. Konagaya, K. Ihara, T. Shimizu, and K. Kimura, "Examination of practicability of communication robot-assisted activity program for elderly people," *Journal of Robotics and Mechatronics*, vol. 23, p. 3, 2011.
- [24] G. Livingston, A. Sommerlad, V. Orgeta, S. G. Costafreda, J. Huntley, D. Ames, C. Ballard, S. Banerjee, A. Burns, and J. Cohen-Mansfield, "Dementia prevention, intervention, and care," *The Lancet*, vol. 390, pp. 2673-2734, 2017.
- [25] C. Anderson-Hanley, P. J. Arciero, A. M. Brickman, J. P. Nimon, N. Okuma, S. C. Westen, M. E. Merz, B. D. Pence, J. A. Woods, and A. F. Kramer, "Exergaming and older adult cognition: a cluster randomized clinical trial," *American journal of preventive medicine*, vol. 42, pp. 109-119, 2012.
- [26] J. Fan, D. Bian, Z. Zheng, L. Beuscher, P. A. Newhouse, L. C. Mion, and N. Sarkar, "A Robotic Coach Architecture for Elder Care (ROCARE) based on multi-user engagement models," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, pp. 1153-1163, 2017.
- [27] J. Fan, L. Beuscher, P. Newhouse, L. C. Mion, and N. Sarkar, "A Collaborative Virtual Game to Support Activity and Social Engagement for Older Adults," in *International Conference on Universal Access in Human-Computer Interaction*, 2018, pp. 192-204.
- [28] J. Fan, J. W. Wade, D. Bian, A. P. Key, Z. E. Warren, L. C. Mion, and N. Sarkar, "A Step towards EEG-based brain computer interface for autism intervention," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, pp. 3767-3770.
- [29] D. M. C. Doerflinger, "Mental status assessment in older adults: Montreal Cognitive Assessment: MoCA Version 7.1 (original version)," *The Clinical Neuropsychologist*, vol. 25, pp. 119-126, 2012.

## CHAPTER 5

### FIELD TESTING OF ROBOT-MEDIATED TRIADIC INTERACTION

#### 5.1 Brief Summary

In CHAPTER 3 and CHAPTER 4, we have designed and developed SAR systems to engage two older adults simultaneously in multimodal activities with the robot. One activity was based on an imitation game named “Simon says” and the other activity was a virtual book sorting activity. Both activities were designed to provide physical, cognitive, and social stimuli to pairs of older adults through triadic interaction. Two user studies were conducted to evaluate system performance and older adults’ acceptance. Despite the promising results from the user studies, the systems were tested in the laboratory setting and older adult participants only interacted with the systems once. In this chapter, we continued the assessment of our SAR systems by conducting a multi-session field study.

The integrated SAR system for triadic interaction, Ro-Tri, consisted of i) a Kinect sensor for gesture recognition, movement mapping, as well as head pose and audio source angle data collection, ii) a Razer Hydra controller for button input detection during “Simon says” activity, iii) a humanoid NAO robot and a 32-inch computer monitor for task presentation, and iv) two empatica E4 wristbands for physiological signals collection. Ro-Tri was tested at two local retirement communities with seven pairs of older adults. Each pair of older adults interacted with Ro-Tri twice a week for three weeks. The primary robot-mediated activities were the “Simon says” activity alternated with the book sorting activity. In addition to subjective data collection, Ro-Tri was able to gather objective interaction data, head pose, vocal sound, and physiological signals in order to automatically evaluate older adults’ activity engagement and social engagement.

This pilot field study lasted 6 months. We had 91.7% attendance over a 100% activity completion rate per session. Survey and questionnaire data were collected from older adult participants and staffs. Results indicated that older adults’ acceptability, tolerance, and interest in the Ro-Tri system as well as the activities were positive and some had slight improvement post-experiment. Results from objective data analysis indicated overall maintained engagement throughout the study. Older adults’ visual attention towards their peers and the system during HRI improved slightly from session one to session six. Participants’ verbal communication improved in the middle of the study but these improvements did not last until the end of the study. Physiological signals analysis demonstrated the ability to monitor older adult’s stress level with 75% accuracy for 36 minutes of data. We believe this is the first work to study and

demonstrate the tolerance and acceptance of older adults in a real world setting and changes of their activity engagement and social interaction over time.

## **5.2 Abstract**

The population in the US is aging rapidly. Evidence based on recent research indicates that there are substantial health benefits in engaging older adults in physical, social, and cognitive activities. Given the lack of healthcare resources, socially assistive robotic (SAR) systems have been developed to administer activity-oriented therapies. Although a growing number of SAR systems have been developed for one-to-one and/or many-to-one interaction with older adults, most systems were tested in the laboratory setting only. In this paper, we present a field study of our SAR system, Ro-Tri, aimed to provide robot-mediated triadic interaction. Ro-Tri was designed to engage older adults in three physically, socially, and cognitively simulating activities over a six-session long interaction. We tested Ro-Tri at two local retirement communities with seven pairs of older adults. In addition to subjective data collection, Ro-Tri is able to gather objective interaction data, head pose, vocal sound, and physiological signals in order to automatically evaluate older adults' activity engagement and social engagement. Results indicate that older adults' visual attention towards their peers during HRI improved slightly from session one to session six, their interest, perception, and engagement in the robot-mediated activities were either maintained or slightly improved. Results also demonstrate the ability of gathered data to assess changes of older adults' engagement and physiological indicators.

## **5.3 Introduction**

The population in the US is aging rapidly as the first batch of baby boomers started turning 65 in 2010. The number of older people is projected to represent nearly 21 percent of the total population in 2030, which is twice as large as the number in 2000 [1]. With aging, many older adults experience chronic health conditions, functional limitations, dementia, and problems with physical functioning, falls, and mental health [1-4]. Dementia, including Alzheimer's disease and other related disorders, is overwhelmingly faced by older adults. One in ten people age 65 and over has Alzheimer's disease in the USA. Dementia impacts communication and interaction ability, impairs judgement, memory, and affect regulation. An additional 15 to 20 percent of older adults have mild cognitive impairment (MCI) and are at higher risk of later developing dementia [5]. The health care costs for older adults with concomitant medical conditions and physical and cognitive impairments are substantial [6, 7]. Informal unpaid caregivers such as family and friends provide 83 percent of the assistance and are under high financial, emotional, and physical burden

[7, 8]. Thus, there is an urgent need for technological strategies that can coexist within resource strained environments to augment the process of effective care for older adults.

Evidence suggest that exercise and physical activity, lifelong learning/cognitive training, and healthy diet may reduce the risk of cognitive decline and dementia. Evidence is growing that social isolation is a risk factor for dementia and social and cognitive engagement may reduce such risks [5, 9, 10]. Physical, cognitive, and social activities have also been shown to improve older adults' physical and psychological well-being and reduce the risks of many health problems such as falls [1, 11, 12]. Although neither pharmacologic nor non-pharmacologic therapies can treat dementia or slow or stop their progression at present, reviews and meta-analyses indicated that cognitive intervention, exercise and physical activity intervention are beneficial to people with Alzheimer's disease and have positive effects on cognitive function [13-15]. Instead of favoring a single intervention, the literature on non-pharmacologic therapies suggests multimodal strategies that tailored to the individual and highlights the importance of social engagement in addition to older adults' physical and mental health [9, 16, 17]. This leads to the development of robotic systems to target the physical, cognitive, and social aspects of older adults.

These therapeutic robotic systems can be categorized into animal robots to provide companionship and address mental illness [18], telepresence robots to facilitate social connections with families and caregivers [19], and socially assistive robotic (SAR) systems to provide activity-oriented therapies such as physical exercise and memory games [20]. SAR systems, including animal robots, are designed specifically for social interactions with capabilities of autonomously detecting and meaningfully responding to older adults' attention and behavior, and thus have significant potential for addressing physical, cognitive, and social conditions. Early studies either used the Wizard of Oz (WoZ) experimental paradigm [21, 22] that requires a human operator to control the robot or used open-loop robotic platforms [23-27] with pre-programmed robotic behaviors. WoZ design places interaction burden on human operator whereas open-loop robotic platforms are limited in their capacity for HRI and lack real-time dynamic adaption based on interaction. More advanced closed-loop robotic systems allow the robot to dynamically alter its interaction based on real-time human interaction. Commercially available robots NAO, RoboPhilo, and Manoi-PF01 have been programmed to instruct older adults and correct their gestures during physical exercise routines [28-31]. A number of researchers have experimented with closed-loop platforms to engage older adults in eating [20, 32], cognitive stimulation [20, 32-34], and chair exercises [35]. The majority of these closed-loop SAR systems are developed for one-to-one HRI to engage older adults in physical and cognitive activities. Research on many-to-one HRI has recently emerged to facilitate group cognitive stimulation, chair exercise, and conversation [24, 27, 34, 36, 37]. These studies have had promising results in engaging



older adults with varying activities. However, most systems were tested in the laboratory setting and older adult participants only interacted with the systems once.

Several clinical trials have been conducted in long term care (LTC) settings involving residents with dementia that examined the effect of a SAR on engagement and various neuropsychiatric symptoms [38]. The most frequently used SAR has been PARO, an animal SAR designed specifically for those with dementia [39]. Clinical trials of PARO have been conducted in LTC settings in Japan, Australia, New Zealand, Norway, and the US [40-45]. Studies varied in sample size (10 to 415), research design (pre-post, cross-over, nonrandomized and randomized clinical trials, cluster randomized trials), intervention design (individual versus group, facilitated versus non-facilitated sessions), length of sessions (10-45 minutes), frequency of sessions (1-3 sessions/week), duration of intervention (1-12 weeks), and outcomes (depression, apathy, quality of life, sleep, agitation, and psychoactive medications). Studies yielded mixed results, but with enough evidence of efficacy for an animal SAR to aid some older adults with various neuropsychiatric symptoms. However, animal robots are limited in their ability to actively engage older adults in cognitive and physical activities since its sole intent is to provide social or emotional connectedness. To the best of our knowledge, clinical trials have not been conducted to evaluate the efficacy of more advanced SAR systems that provide activity-oriented therapies.

A few researchers have evaluated the performance and user acceptance of more advanced SAR systems in the field. Robot Brian 2.1 was placed at a LTC facility for two days and interacted with 40 older adults to play a memory card game or monitor a meal-eating activity [20]. Robot Tangy scheduled and played Bingo game with seven residents at a LTC facility [46]. Each resident participated in at least two sessions out of the six total group sessions. Field trial of robot Matilda was conducted with 70 residents from three residential care facilities over a three-day period [36]. A robotic exercise tutor was tested with six residents in a nursing home for a single session HRI and tested with 12 visitors of a day care center for multi-session HRI (1 – 5 sessions, mean: 2.58 sessions) [47]. Overall, the majority of the participants in these studies were engaged and complied with the SAR systems. However, the participants had very limited exposure to the SAR systems with the majority of them only interacted with the system once. In addition, none of these systems were developed to facilitate human-human interaction (HHI) and social interaction among older adults were either not observed or not discussed in field studies conducted in group setting.

In [48] and CHAPTER 4, we have designed and developed SAR systems to engage older adults simultaneously in physical, cognitive, and social activities with the robot. One activity was based on an imitation game named “Simon says”, where each older adult and the robot took turns to direct a gesture and expected that others would follow only if the gesture was introduced with utterance “Simon says”. Another activity was a virtual book sorting task with two triadic interaction modes, take turns interaction and

simultaneous interaction. These systems were tested with pairs of older adults in the laboratory setting. Collectively, results from two user studies indicated the potential for SAR-based interaction to involve more than one older adult, to administer multimodal activities with the aid of the robot, and to quantitatively measure older adults' social interaction and activity engagement. In order to assess the feasibility, acceptability, tolerance, and potential impact of the robot-mediated triadic interaction with paired older adults, we conducted a pilot field study with 7 pairs of older adults residing at two local retirement communities. Each pair of older adults received the interaction for 6 sessions over a period of 3 – 4 weeks. The main focus of this work is to determine i) whether robot-mediated triadic interaction is feasible, well-tolerated, and accepted by older adults; ii) whether older adults remain engaged over time and attend all the sessions; iii) the robot's ability to encourage communication and social engagement between the two older adults; and iv) the feasibility of gathering data and demonstrating changes of older adults' task performance, physiological indicators, and involvement/engagement in activities. To the best of our knowledge, this is the first work to study and demonstrate the tolerance and acceptance of older adults in a real world setting and changes of their activity engagement and social interaction over time. The rest of the paper is structured as follows. Section 5.4 describes the SAR system, field study, and data collection and analysis methods. Section 5.5 presents the objective and subjective data analysis results. Finally, the results and implications are discussed in Section 5.6.

## **5.4 Method**

### **5.4.1 Robotic system**

The SAR system for triadic interaction (Ro-Tri) was the combination of our previous SAR systems to administer the “Simon says” activity and the book sorting activity. Due to the fact that the two SAR systems were not implemented in the same development environment and they used different hardware versions, we chose not to unify the two systems as one integrated system. We did not intentionally change development environment but conformed to the requirements of upgraded equipment. However, the two systems have similar architectures that consisted of modules and submodules described in ROCARE, a multi-user SAR architecture we proposed in [49]. This property enabled us to combine the two system architectures and to integrate the two systems to a certain degree.

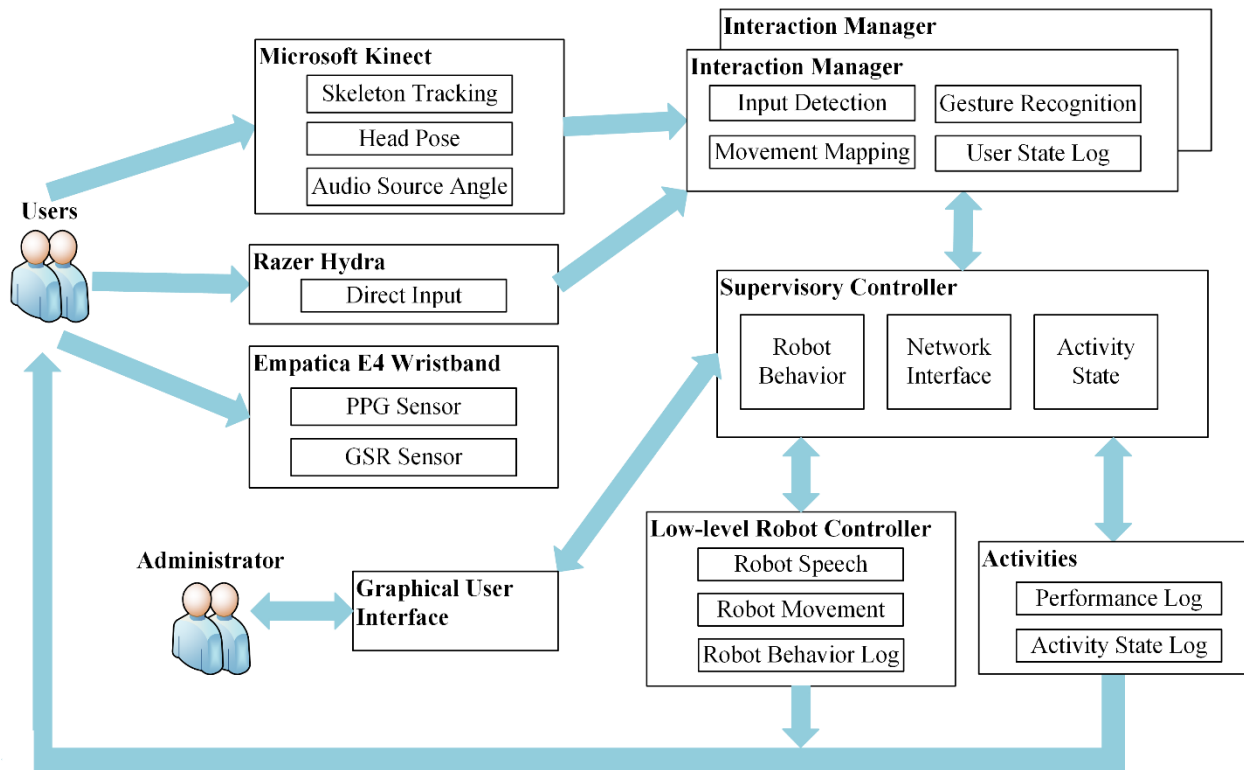


Figure 5-1. Ro-Tri Architecture

The integrated Ro-Tri architecture is illustrated in Figure 5-1. The sensing module was composed of i) a Microsoft Kinect for Windows RGB-D sensor for online gesture recognition and movement mapping based on skeleton data, as well as offline evaluation of activity engagement and social interaction based on head pose and audio source angle data; ii) a Razer Hydra controller for input detection during “Simon says” activity; and iii) two empatica E4 wristbands for offline analysis of physiological indicators. The triadic interaction was mediated by a humanoid NAO robot via robot speech and gestures. The supervisory controller module communicated with the interaction managers, the activities displayed on a 32-inch computer monitor or administered directly by the robot, and the low-level robot controller for autonomous closed-loop interaction. The graphical user interface module enabled an administrator to initiate and monitor the interaction. The majority of these modules were exactly the same as in the previous SAR systems. The modifications applied to integrate these two systems included: i) we removed the EEG acquisition module due to the relatively time-consuming process to put on the sensors for two older adults. As an alternative, we included the E4 wristbands to collect physiological signals; ii) we extended the interaction duration of the “Simon says” activity to match with the interaction duration of the book sorting

task; and iii) we modified the quantitative data acquisition modules for the “Simon says” activity to match with that of the book sorting task.

Ro-Tri, as shown in Figure 5-2, was capable of administering four activities, which were “Simon says”, book sorting (take turns), book sorting (simultaneous), and book sorting (post-test). For the sake of completion, we describe the behavior of Ro-Tri for each activity. Details of the system development and preliminary user study results can be found in our previous papers. In “Simon says” activity, we used Kinect v1 for gesture recognition and the computer monitor was turned off. There were five rounds of group interaction. The first round was introduction, during which the robot and older adults took turns to introduce their names and then meet each other. The second to fourth round were “Simon says” play. In each round, the robot first acted as a leader to demonstrate an arm movement, first time with “Simon says” that required older adults to copy the movement and second time without “Simon says”. Then, the robot asked one older adult to play as the leader. The robot and the other older adult would be the followers. This round ended with the other older adult also had the chance to play as the leader. The robot was programmed to demonstrate and recognize three gestures, which were wave, raise both arms up, and extend arms to the side. When older adults demonstrated a gesture, the robot had the ability to mirror their upper arm movements. The last round was when the robot thanked the older adults and asked them to wave goodbye to each other.

Book sorting activity used Kinect v2 for motion-based interaction with a virtual reality (VR)-based book sorting game displayed on the computer monitor. The VR-based book sorting game consisted of different colored books and color matched bins to put books into. Each older adult had a hand cursor displayed on the monitor that they could manipulate through large range arm movements and open/close hand gestures. For example, when older adults moved their arms to the left, the hand cursor would move to the left of the monitor until it reached the left boundary. When older adults’ hand cursors overlapped with books, they could grab the books by closing their hands. We defined rules in the game to reward collaboration behaviors occurred during interaction. In take turns and simultaneous interaction, collaboration happened when older adults helped each other by moving books closer to each other’s bins. Whereas in post-test, collaboration happened when older adults moved the same book in the same direction. In take turns interaction, there was only one hand cursor displayed on the monitor, and older adults were required to wait for their peers to finish before they could control the hand cursor. Whereas in simultaneous interaction, two hand cursors were allowed and older adults could play at the same time. The robot facilitated the older adults in take turns and simultaneous interaction with the purpose of maintaining and enhancing task engagement and HHI. This was realized by continuously evaluating older adults’ interaction and providing feedback to engage them in motion-based interaction, encourage them to help each other,

and celebrate their accomplishment in the game. In the post-test, older adults were not told to move the same book together. Through social interaction, we expected them to explore different ways to interact with the system and gradually figure out that they needed to collaborate to move books. The robot would provide a hint half way through the interaction if older adults were not able to move books at all.

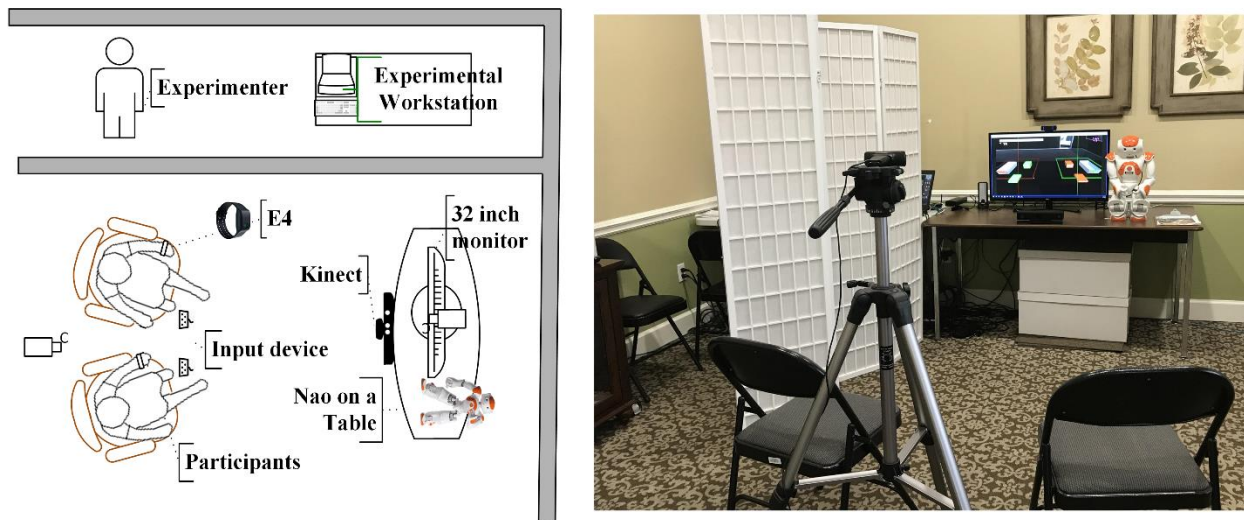


Figure 5-2. Ro-Tri System Overview and Setup at a Retirement Community

#### 5.4.2 Field study

This study was approved by the Vanderbilt University Institutional Review Board. The system was placed at local retirement communities and used by older residents. The eligibility criteria for participant recruitment included: i) age 70 or older; ii) ability to hear as screened by the Whisper Test. Participants may use hearing aids; iii) ability to see as screened by ability to read newspaper print. Participants may use eyeglasses; iv) ability to move arms as screened by the ability to raise arms up, forward and to the side; and v) able to cognitively participates in various robotic activities designed for them. The experimental setup and materials are shown in Figure 5-2. Participants sat in front of and facing the system. NAO was positioned by the side of the computer monitor. The Kinect was placed on the edge of the table facing two participants. An administrator operated experimental workstation in a separated space. The primary robot-mediated activities for the paired older adults were the book sorting activity alternated with the “Simon says” activity (Table 5-1). Each older adult pair interacted with Ro-Tri twice per week for three weeks within a month. Before the triadic interaction, each participant went through an orientation to get familiar with the virtual book sorting activity as well as the robot movements and speech. The estimated interaction duration only included the time needed to interact with Ro-Tri. The whole session also involved putting on sensors, calibration and baseline data recording, adjusting robot speaking volume, adjusting program

parameters, as well as collecting subjective data from older adults and thus lasted approximately 40 minutes for each pair of older adults.

Table 5-1. Experimental Protocol

Session (Week)	Activity Description	Estimated Duration	Comment
Orientation	one-to-one HRI & task orientation	9 min	Older adults get familiar with interacting with the system and the robot.
Session 1 (Week 1)	“Simon says”	9 min	--
Session 2 (Week 1)	book sorting (take turns)	9 min	Allow older adults to practice in the virtual environment for 3-5 min before HRI.
Session 3 (Week 2)	book sorting (simultaneous) & book sorting (post-test)	12 min	Allow older adults to practice in the virtual environment for 3-5 min before HRI.
Session 4 (Week 2)	“Simon says”	9 min	Repeat Session 1
Session 5 (Week 3)	book sorting (take turns)	9 min	Repeat Session 2
Session 6 (Week 3)	book sorting (simultaneous) & book sorting (post-test)	12 min	Repeat Session 3

We conducted field study first at Sycamores Terrace Retirement Community with 9 older adults and then at Elmcroft Senior Living with 6 older adults. At Sycamores Terrace Retirement Community, Ro-Tri was set up in an apartment. Participants interacted with Ro-Tri in the living room whereas the administrator operated the experimental workstation in the bedroom. At Elmcroft Senior Living, Ro-Tri was set up in the corner of a library with a room divider to separate the experimental workstation from participants. A total number of 14 older adults (7 pairs, mean age: 82.7, 3 had normal cognition, 10 had MCI, and 1 had Alzheimer’s dementia) completed the field study. One older adult dropped out after second session due to her hearing aid issue and her peer was paired with another older adult and restarted from session one.

At the start of each session, we first put E4 sensors on participants’ non-dominant wrists and recorded three minutes of baseline physiological responses while the participants were asked to sit quietly. We then reminded them how to interact with the system. In “Simon says” sessions, we told them only arm movements were recognized by the robot and reminded them to pull the trigger button of the Razer Hydra controller after they answered robot’s questions. In book sorting sessions, we asked them to practice moving their hand cursors and grabbing books for a few minutes. Practice was followed by a short calibration which recorded Kinect’s head pose angles when we asked older adults to look at the robot, the computer monitor, and their peers as well as Kinect’s sound source angles when we asked each older adult to read a sentence. Administrators then started the interaction and stayed out of sight of the participants during the interaction. Finally, participants filled out a post experiment evaluation questionnaire at the end of session 2, 4, and 6.

### 5.4.3 Data collection and analysis

Two types of data were gathered during the study, which were objective data logged automatically by Ro-Tri and subjective data filled out by participants and caregivers. Subjective data included surveys for participants' acceptance of the system and Visual Analog Scale (VAS) for caregivers' opinion about participants. Prior to implementation and conclusion of the study, participants completed the Robot Acceptance Scale (RAS, 7-point scale, 1 most positive to 7 most negative response) we have developed for our previous robot studies [50], and staff completed a VAS (0 – 10 continuous scale, 0 most negative to 10 most positive response) for assessment of the extent to which participants interacted with others and were interested in the robot sessions. At the end of each week, participants completed a post experiment evaluation questionnaire (7-point scale, 1 most negative to 7 most positive response) that provided opinions about the activities and robot sessions.

Objective data collected were participants' interaction data and activity states, participants' head pose angles, Kinect's sound source angles as an indicator of sound source direction, participants' physiological responses from E4 sensor, and robot's behaviors. Interaction data logged participants' interaction with the book sorting task as well as their upper body skeleton position data. From interaction data and activity states, we computed an effort metric representing the amount of effort exerted by the participants during HRI. For book sorting tasks, the effort was the amount of book movements to collect one's own book or to help others. For "Simon says" activity, the effort was the accumulated elbow and wrist movements. Participants' head pose yaw angles served as a coarse estimation of their gaze directions. The head pose yaw angles were zero when participants looked straight ahead, decreased when they looked to the right, and increased when they looked to the left. From the calibration data, which logged participants' head poses when they looked at the computer monitor, the robot, and their peers, we calculated head pose yaw angle ranges for head towards the robot, head towards the computer monitor, and head towards the other person (Figure 5-3). These ranges allowed us to compute automatically the amount of times older adults' paying visual attention to the computer monitor or the robot, as well as the amount of times and the number of times older adults moved their heads towards their peers. We considered visual attention to the system as activity engagement and visual attention to the other older adult as social engagement.

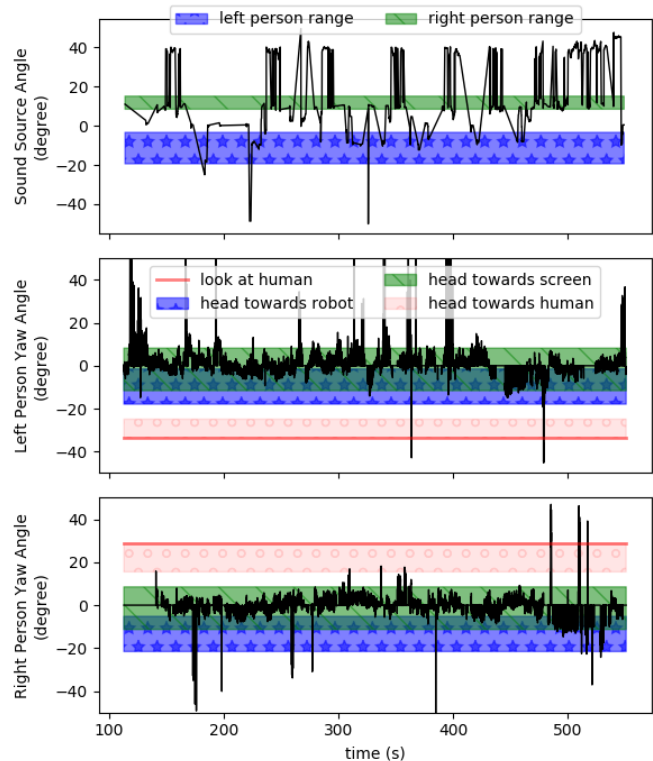


Figure 5-3. Raw Sound Source Angle and Head Pose Yaw Angle Data for One Session

To compute activity engagement based on head pose yaw angles, first, the ranges of head towards computer monitor and robot and the thresholds for head towards human were used to segment raw head pose data into intervals of data that belong to activity engagement (head towards computer monitor or robot), social engagement, or neither. Then, the intervals belonging to activity engagement were summed together to calculate the total activity engagement duration. For social engagement, we first generated candidates of start timestamps when older adults potentially initiated a looking behavior. These candidates were selected from the intervals belonging to social engagement. In order to reduce accidental count of head turns due to noisy data, we set a 1s threshold so that the start time of the next head turn must be 1s ahead of the end time of the previous head turn. The end timestamp for a selected candidate was calculated by merging the intervals associated with the candidate and outputting the end time of the merged interval. Each candidate represents a potential head turn. Some of these candidates were noises that have very short durations. Some of these candidates were generated because older adults' hands were in front of their faces. Some of these candidates were the results of head pose data interpolation. All the candidates were passed through three thresholds to filter out the abovementioned artifacts. From the remaining candidates, we calculated the social engagement duration and the number of times older adults looked towards their peers.



The sound source angles data were used to estimate the start and end of vocal sounds made by older adults. From calibration data, we were able to compute ranges of sound source angles that capture each older adult’s vocal sounds. As shown in Figure 5-3, the green bands indicate when the right person was talking and the blue bands indicate when the left person was talking in one HRI session. To compute automatically the amount of times older adults were talking and the number of times they spoke, first, we segmented the raw sound source angles into intervals of data that belonged to the left speaker, right speaker, or neither based on the ranges and confidence levels of the detection algorithm. Second, the start times and end times of these intervals were mapped to their largest previous and smallest following integers, respectively, by applying the floor and ceiling functions. After this mapping, some intervals might overlap. We then merged all the overlapped intervals and finally summed the duration of these intervals to calculate the total amount of time older adults making vocal sounds during triadic HRI. We also computed the number of times they were speaking as the count of these intervals after merging.

The abovementioned algorithms for automatically computing the amount of times and the number of times older adults talked and looked towards the other person were validated using data recorded during previous laboratory test. In our previous laboratory experiment, paired older adults performed book sorting tasks (take turns and simultaneous) under the guidance of robot. A trained research assistant manually analyzed video and audio recordings and logged the start and end timestamps for each talking and looking behavior as the ground truth. The start and end timestamps automatically generated by the algorithms were validated against the ground truth. We validated the algorithms based on data from 8 older adults. The validation results are shown in Table 5-2. In general, head pose analysis algorithm could detect with high accuracy the amount of times and the number of times older adults looked towards their peers. The start time deviation for correctly detected looks has a mean value of 0.25s and a standard deviation of 0.14s. The end time deviation for correctly detected looks has a mean value of 0.30s and a standard deviation of 0.21s. The sound source angle analysis algorithm could detect with high accuracy the number of times older adults spoke. For the duration of speaking, the algorithm has high precision but many missed detections. Therefore, the speaking duration was excluded in our field study data analysis.

Table 5-2. Validation Results of Automatic Evaluation Algorithms

<b>Data Type</b>	<b>Measure</b>	<b>Precision</b>	<b>Recall</b>
Head Pose	Amount of times looking towards peers	98.30%	91.52%
	Number of times looking towards peers	95.65%	86.27%
Sound Source Angle	Amount of times talking	99.40%	65.41%
	Number of times talking	97.92%	87.04%

E4 sensor was used to record peripheral physiological data including photoplethysmogram (PPG) and electrodermal activities (EDA). The sampling rates for PPG and EDA were 64Hz and 4Hz, respectively. The data were examined and multiple features were extracted (Table 5-3). Heart rate (HR) was computed by detecting peaks in the PPG signal. HR reflects emotional activity. Generally, it has been used to differentiate between positive and negative emotions. Heart rate variability (HRV) measures the specific changes in time (or variability) between successive heart beats. HRV refers to the oscillation of the interval between consecutive heartbeats. It has been used as an indication of mental effort and stress in adults [51]. EDA provides a measure of the resistance of the skin. This resistance decreases due to an increase of sudation, which usually occurs when one is experiencing emotions such as stress or surprise. Tonic and phasic components of EDA were decomposed separately from the original signal [52]. The tonic component is the baseline level of EDA and is generally referred to as skin conductance level (SCL). The phasic component is the part of the signal that changes when stimuli are presented and is known as skin conductance response (SCR). Lang et al. discovered that the mean value of the SCR is related to the level of arousal [53]. EDA is a strong indicator of affective arousal in general [52]. Gjoreski et al. has used skin temperature data from a E4 wristband to predict stress level [54]. From the physiological data, we were interested to see whether it was possible to detect times when the participants were stressed and when they were relatively at ease during HRI. The three minute baseline data were used to remove feature variations due to time and individual difference. Specifically, the heart rate, heart rate variability, mean SCL, mean amplitude of SCR, and mean skin temperature features were subtracted by their respective baseline values and divided by their respective baseline standard deviation values. For the remaining features, the baseline values were subtracted from the features.

Table 5-3. Peripheral Physiological Features

<b>Physiological Signal</b>	<b>Features</b>	<b>Unit of Measurement</b>
PPG	Heart rate	Beats/min
	Heart rate variability	ms
EDA	Mean SCL	$\mu\text{S}$
	Standard deviation of SCL	$\mu\text{S}$
	Mean amplitude of SCR	$\mu\text{S}$
	Standard deviation of SCR	$\mu\text{S}$
	Maximum amplitude of SCR	$\mu\text{S}$
	Rate of SCR	Response peaks/s
Infrared Thermopile	Mean skin temperature	$^{\circ}\text{C}$
	Standard deviation of skin temperature	$^{\circ}\text{C}$

## 5.5 Results

### 5.5.1 Objective data analysis results

The system worked as designed. Fourteen participants completed all 6 sessions. One participant dropped out after completion of session 2 due to issues with her hearing aids. For the post-test task, older adults were able to figure out the unknown collaborative rule and move yellow books together through communication with their peers for 10 out of 14 sessions. Robot provided hints to help them for the rest of the 4 sessions. Table 5-4 lists participants' engagement across 6 sessions as measured by their interaction effort, head pose, and sound source angle. On average, participants spent 77.7% of the time looking at Ro-Tri and 2.3% of the time looking towards their peers. The number of times looking towards peers and the number of times talking across 6 sessions were 0.41 times per minute and 3.72 times per minute, respectively. The duration of each looking behavior towards their peers had an average value of 2.94 second.

Table 5-4. Participants' Engagement across Six Sessions

Data Type		Session1 M (SD)	Session2 M (SD)	Session3 M (SD)	Session4 M (SD)	Session5 M (SD)	Session6 M (SD)
Interaction Data	Effort / min	1.08 (0.77)	13.19 (5.45)	14.64 (8.18)	1.08 (0.73)	13.33 (6.34)	13.74 (5.12)
Head Pose Activity Engagement	Duration (Percentage)	71.0% (14.6%)	87.6% (6.9%)	76.1% (15.5%)	70.7% (19.2%)	84.1% (9.8%)	76.4% (14.1%)
	Duration (Percentage)	1.3% (1.5%)	1.7% (3.0%)	3.1% (6.0%)	3.2% (4.7%)	2.0% (3.7%)	2.4% (3.1%)
Head Pose Social Engagement	Count / min	0.29 (0.32)	0.48 (0.51)	0.50 (0.56)	0.35 (0.41)	0.37 (0.41)	0.47 (0.54)
	Duration / count	2.82 (1.75)	1.71 (0.87)	2.57 (2.46)	4.89 (3.45)	2.61 (1.88)	3.04 (1.33)
Sound Source Angle	Count / min	2.48 (2.29)	3.67 (4.17)	5.35 (5.64)	3.41 (1.71)	3.80 (4.52)	3.59 (3.59)

Since participants' interaction effort, visual attention, and communication varied for different activities, we normalized the engagement results in order to compare results and demonstrate changes over 6 sessions. For each activity, "Simon says", book sorting take turns, and book sorting simultaneous, we computed the best engagement values achievable for this target population by taking the average of the top three values for that activity. The worst engagement values were derived based on the nature of the

engagement measure and the activity. For effort, visual attention towards peers, and verbal communication, the worst values were zero. Whereas for visual attention towards the system, the worst value was the head pose range towards the system divided by 180 assuming that participants looked at different directions randomly. We then normalized the engagement results by first subtracting the worst engagement value, then divided by the absolute difference between the best and worst engagement values. After normalization, the higher the value, the better the engagement results.

Figure 5-4 demonstrates changes of interaction effort, visual attention, and verbal communication over 6 sessions. The group results were the mean values of the 14 participants. In addition to group results, we plotted changes of engagement for some individual older adults as examples. Older adults' interaction effort maintained throughout the HRI sessions with very slight improvement towards the end, 2.9% at session 6. Eight out of 14 participants' effort increased from session 1 to session 6. For head pose data, participants' activity engagement represented by percentage of time they looked at the system increased slightly, by 7.2% at session 6. The change of visual attention over 6 sessions was very different for each individual, as illustrated by two participants' results (S211 and S305). From session 1 to session 6, nine participants paid more visual attention to Ro-Tri and five of them paid less attention. Seven out of 8 participants who paid more attention to the system also had increased interaction effort.

The percentage of time participants looked at their peers continued increasing from session 1 to session 4. Eventually, participants' visual attention towards their peers increased by 4.7%, which was slightly less than the increase of their visual attention towards the system. Seven participants paid more visual attention to their peers from session 1 to session 6. Together with visual attention to the system, only 2 participants had decreased visual attention. The rest of the participants either paid more visual attention to both HRI and HHI (4 out of 12) or paid more attention to the system and less attention to their peers or vice versa. The number of times participants looked towards their peers also increased slightly, by 4.7% at session 6. Overall, participants looked towards their peers at a frequency similar to session 1 during the experiment. However, they spent longer duration for each looking behavior, increased by 8.8%. Finally, for verbal communication results, participants talked more during week 2 as compared to week 1. During week 3, their verbal communication results reduced, even fell below to that of week 1.

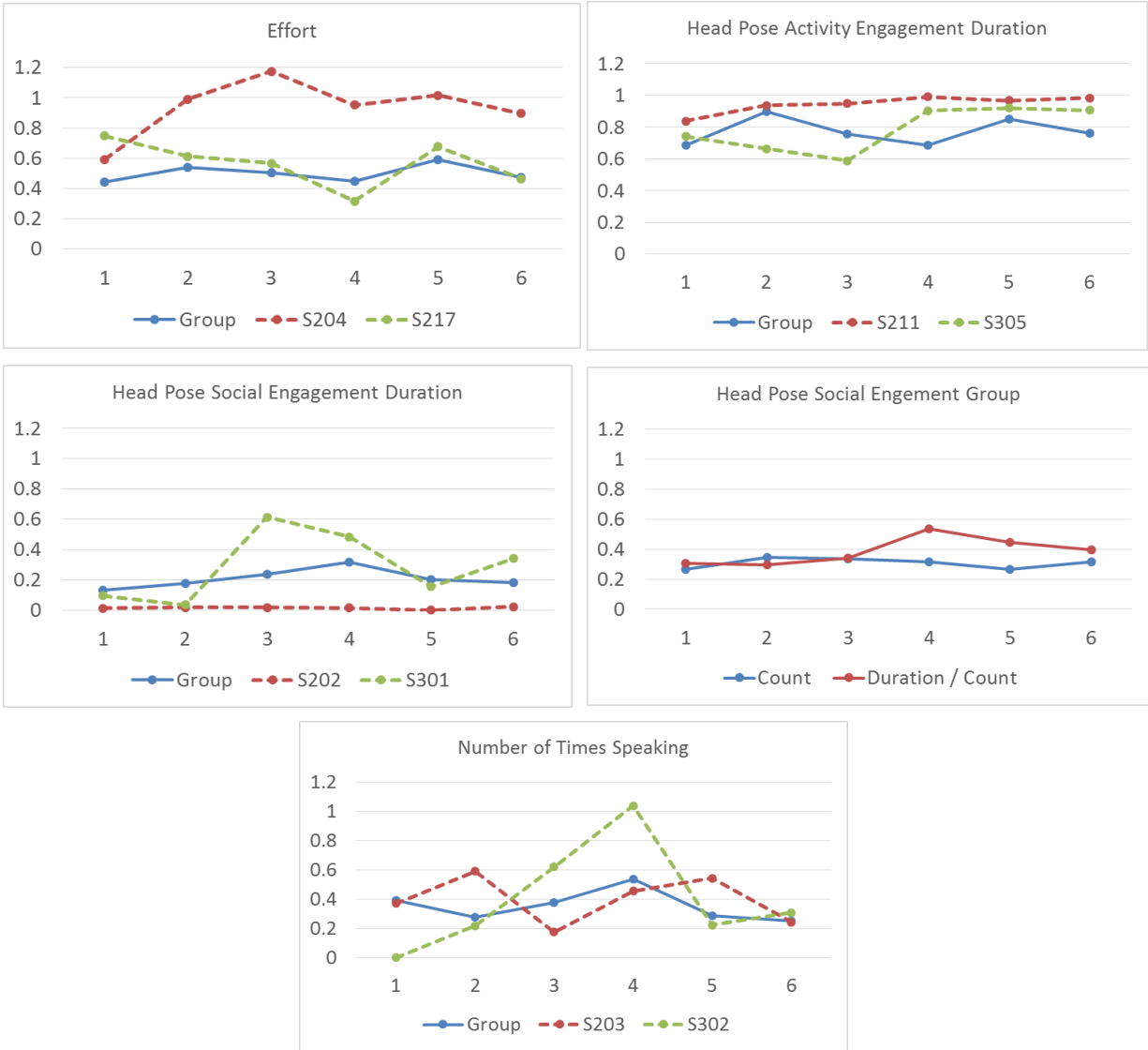


Figure 5-4. Changes of Engagement Over Six Sessions

By observing the video recordings of the experiment, we selected 14 instances of data where the participants were stressed and 14 instances where the participants were relatively calm. Each instance was one and half minute in duration and labeled by a research assistant, either as “calm” or “stressed”. Waikato Environment for Knowledge Analysis (WEKA) was used for feature selection and model training. The wrapper subset evaluation method using the best attribute technique (forward direction) was used to select the best features and four machine learning algorithms were used to predict the stress level. The machine learning algorithms were evaluated with five-fold cross validation. The best performing features selected were mean SCL, mean SCR, mean heart rate variability, and mean skin temperature. The machine learning algorithms applied as well as the corresponding classification results are shown in Table 5-5.

Table 5-5. Classification Results

<b>Machine Learning Algorithm</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F-1 score</b>
Random forest	75.0%	0.75	0.75	0.75
Decision stump	71.4%	0.76	0.71	0.70
Decision tree (J48)	71.4%	0.71	0.71	0.71
Logistic regression	67.8%	0.74	0.68	0.66

### 5.5.2 Subjective data analysis results

The RAS survey was conducted to determine participants’ acceptance and anticipated use of the robotic system based on performance expectancy, effort expectancy, and attitude towards using the system. All participants completed the pre RAS and 13 of them completed post RAS. Participants’ perceptions became more positive for all the subscales and RAS after the experiment (Table 5-6). Wilcoxon signed-rank test results are shown in the table, including the standard score of the Wilcoxon signed ranks, p value, and effect size. The improvements of the perceptions were not statistically significant. Effort expectancy subscale, attitude subscale, and RAS were more positive with medium effect size. VAS was completed by caregivers or staffs that were familiar with the participants. The five questions and their results are shown in Table 5-7. After six HRI sessions, staffs’ ratings on participants’ social interaction during daily activity improved by 6.2%. Participants were observed to be more interested with Ro-Tri, 8.2% improvement on anticipation of robot session and 2.3% decrease on complain about robot session. Participants’ engagement on daily activities decreased by 1.6%. None of these changes was statistically significant using the Wilcoxon signed rank tests.

Table 5-6. RAS Results

	<b>Pre<sup>a</sup></b>	<b>Post<sup>a</sup></b>	<b>Z</b>	<b>p</b>	<b>r</b>
	<b>M (SD)</b>	<b>M (SD)</b>			
Performance Expectancy	2.84 (0.57)	2.69 (0.78)	0.36	0.749	0.07
Effort Expectancy	2.83 (0.77)	2.52 (0.70)	1.55	0.133	0.30
Attitude	2.57 (0.55)	2.12 (0.66)	1.81	0.075	0.35
RAS	2.70 (0.57)	2.36 (0.67)	1.75	0.084	0.34

<sup>a</sup>Lower values are more positive, 7-point scale

Table 5-7. VAS Results

Question	Pre <sup>a</sup> M (SD)	Post <sup>a</sup> M (SD)	Change	Z	p	r
How social would you rate the participant?	7.97 (2.05)	7.97 (1.69)	0%	--	--	--
To what extent would you say the participant likes to come out and do activities?	7.21 (3.05)	7.05 (2.91)	-1.6%	0.71	0.518	0.14
To what extent would you say the participant likes to talk to other residents, staffs, or family?	8.08 (1.85)	8.71 (1.44)	6.2%	1.54	0.132	0.29
To what extent would you say the participant looks forward to attend the robot sessions	8.62 (1.42)	9.44 (0.71)	8.2%	1.68	0.102	0.36
To what extent would you say you observed the participant complained about the robot sessions?	9.41 (1.33)	9.64 (0.65)	2.3%	0.31	0.844	0.07

<sup>a</sup>Higher values are more positive, 0 - 10 continuous scale

Table 5-8. Post Experiment Evaluation Results

	Week 1 <sup>a</sup> M (SD)	Week 2 <sup>a</sup> M (SD)	Week 3 <sup>a</sup> M (SD)
Interest on robot session	6.33 (0.94)	6.33 (1.26)	6.52 (0.61)
Interest on triadic interaction	6.19 (1.03)	5.88 (1.70)	6.33 (0.91)
Acceptability of robot	6.21 (1.01)	6.33 (0.81)	6.54 (0.55)
Acceptability of activity	6.15 (0.95)	6.40 (0.83)	5.80 (1.10)
Interest on activity	6.19 (1.01)	6.30 (0.78)	5.96 (1.13)

<sup>a</sup>Higher values are more positive, 7-point scale

Post experiment evaluation gathered participants' interests and acceptability on robot sessions after each week of HRI. Examples of questions being asked were "Enjoyed attending the robot sessions" (interest on robot session), "Looked forward to interact with another residents for the robot sessions" (interest on triadic interaction), "The robot was able to keep your attention" (acceptability of robot), "Doing the book sorting activity for future studies" (acceptability of activity), and "How interesting or boring were the Simon says activity" (interest on activity). In general, participants' interests and acceptability on the robot, triadic interaction with peers, and the activity were positive and maintained over 3 weeks (Table 5-8).

## 5.6 Discussion and Conclusion

Due to limited work reported in the literature on testing SAR systems designed to engage older adults in physical and cognitive activities in real world setting, we conducted a field study of SAR systems for triadic interaction that we developed previously to fill the gap. Our system, Ro-Tri, could involve two older adults in three physical, cognitive, and social stimulating activities, which were “Simon says” activity and book sorting activity with take turns and simultaneous interaction mode. Pairs of older adults residing in local retirement community interacted with Ro-Tri for 6 sessions over 3-4 weeks. Both objective and subjective data were collected and analyzed to assess the feasibility, acceptability, tolerance, and potential impact of robot-mediated triadic interaction with paired older adults. As a feasibility study, our goal is to determine the extent to which older adults’ interest and participation in the robot-mediated activities are maintained, demonstrate engagement and social interaction during robot-mediated triadic interaction, and gather data and demonstrate changes of older adults’ engagement and physiological indicators.

The field study lasted 6 months, including the time taken to find a second site and setup system at retirement communities. Fifteen participants were recruited to take part in the study. Eight participants (4 pairs) completed all 6 sessions at the first retirement community and 6 participants (3 pairs) completed the study at the second retirement community. We had 91.7% attendance over a 100% activity completion rate per session. Older adults’ acceptability, tolerance, and interest in Ro-Tri system as well as the activities were studied by collecting survey and questionnaire from older adult participants and staffs. Results indicate that participants’ perceptions on Ro-Tri were more positive after the experiment on all subscales and RAS, their interest and acceptability were high for both the robot and the activity, and they enjoyed interacting with another residents for the robot sessions. RAS survey were sensitive to change from pre- to post-experiment. Although none of the improvements was statistically significant, some had medium effect size which suggests more pairs of participants or longer study might lead to better results.

Ro-Tri logged different objective data during HRI to evaluate older adults’ engagement in terms of interaction effort, visual attention to the system and another older adult, and verbal communication during HRI. In general, participants’ engagement maintained throughout the study. There were slight improvements for visual attention towards the system and peers from session 1 to session 6. The percentage of time they spent looking towards their peers increased by 4.7% and the duration of each looking behavior increased by 8.8%. The percentage of time they spent looking at the system increased by 7.2%. Participants’ verbal communication improved in the middle of the study but these improvements did not last until the end of the study. They may talked less as they become more familiar with their peers and the robot-mediated tasks. On average, participants spent 77.7% of the time paying visual attention to the system (70.9% for



“Simon says”, 85.9% for book sorting take turns, and 76.3% for book sorting simultaneous). For social interaction with their peers, participants looked towards their peers 0.41 times per minute and talked 3.72 times per minute.

E4 sensors were used to collect physiological signals from older adults. It is easy to apply and none of the participants complained about wearing a wristband. We extracted 10 features from the physiological data to estimate the stress level of participants during HRI. The ability to monitor the stress level will enable a future stress-sensitive robotic system. This way Ro-Tri could provide more personalized and effective feedback to engage older adults in activity-oriented therapies. The classification results indicated a 75% classification accuracy for 36 minutes of data. The decision forest algorithm had the highest accuracy among the four algorithms tested.

To the best of our knowledge, this is the first work to study and demonstrate the tolerance and acceptance of older adults in a real world setting and changes of their activity engagement and social interaction over time. We present the feasibility of Ro-Tri to engage older adults in HRI as well as HHI over time in a real world setting based on objective and subjective data analysis results. The uniqueness of this platform relies on its ability to simultaneously involve two older adults with mechanisms to foster HHI through HRI, and its ability to gather objective interaction data, head pose, vocal sound, and physiological signals and automatically evaluate activity engagement and social engagement. Findings from the study indicated slightly improved social interaction during HRI in terms of visual attention towards other older adults and maintained or slightly improved interests, perceptions on the system, and engagement in HRI.

The current work is limited in several ways. First, the sample size and interaction duration are not large enough to provide evidence on the efficacy or impact of Ro-Tri on older adults’ activity and/or social engagement in daily life. Although 6 sessions were a relatively longer exposure to the SAR system as compared to the existing literature, more exposure will provide more information and stronger results on how older adults’ engagement changes over time. Second, the current results focus on the changes of group engagement. As seen in Figure 4, each individual older adult’s interaction effort, visual attention, and verbal communication over 6 sessions changes differently. Third, we observed situations where one older adult performed very well whereas the other performed poorly. Some older adults were sensitive to their performance as compared to their peers and this might change their response to the system. In the future, we will add more robot behaviors to help reduce the gap between two older adults’ task performance. We will also conduct more in-depth analysis of the data, including analysis of experimental videos and looking into each individual’s change of engagement and physiological response. Finally, we will continue the development of more robot-mediated activities and testing of SAR systems based on results and knowledge gained from this field study.

## REFERENCES

- [1] Federal Interagency Forum on Aging-Related Statistics, "Older Americans 2016: Key Indicators of Well-Being," *Federal Interagency Forum on Aging-Related Statistics*, 2016.
- [2] US Department of Health and Human Services, "Multiple chronic conditions—a strategic framework: optimum health and quality of life for individuals with multiple chronic conditions," *Washington, DC: US Department of Health and Human Services*, vol. 2, 2010.
- [3] A. Fiske, J. L. Wetherell, and M. Gatz, "Depression in older adults," *Annual review of clinical psychology*, vol. 5, pp. 363-389, 2009.
- [4] S. M. Friedman and D. A. Mendelson, "Epidemiology of fragility fractures," *Clinics in geriatric medicine*, vol. 30, pp. 175-181, 2014.
- [5] Alzheimer's Association, "2018 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 14, pp. 367-429, 2018.
- [6] Centers for Disease Control and Prevention, "Chronic diseases: The leading causes of death and disability in the United States," *Chronic Disease Overview*, 2015.
- [7] Alzheimer's Association, "2017 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 13, pp. 325-373, 2017.
- [8] B. C. Spillman, "Why do elders receiving informal home care transition to long stay nursing home residency," *Report to Department of Health and Human Services, Assistant Secretary for Planning and Evaluation, Office of Disability, Aging and Long-Term Care Policy, Contract HHSP23337033T*, 2014.
- [9] G. Livingston, A. Sommerlad, V. Orgeta, S. G. Costafreda, J. Huntley, D. Ames, C. Ballard, S. Banerjee, A. Burns, and J. Cohen-Mansfield, "Dementia prevention, intervention, and care," *The Lancet*, vol. 390, pp. 2673-2734, 2017.
- [10] M.-V. Zunzunegui, B. E. Alvarado, T. Del Ser, and A. Otero, "Social networks, social integration, and social engagement determine cognitive decline in community-dwelling Spanish older adults," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 58, pp. S93-S100, 2003.
- [11] W. J. Chodzko-Zajko, D. N. Proctor, M. A. F. Singh, C. T. Minson, C. R. Nigg, G. J. Salem, and J. S. Skinner, "Exercise and physical activity for older adults," *Medicine & science in sports & exercise*, vol. 41, pp. 1510-1530, 2009.
- [12] M. Cattan, M. White, J. Bond, and A. Learmouth, "Preventing social isolation and loneliness among older people: a systematic review of health promotion interventions," *Ageing & Society*, vol. 25, pp. 41-67, 2005.
- [13] C. Groot, A. Hooghiemstra, P. Raijmakers, B. Van Berckel, P. Scheltens, E. Scherder, W. Van der Flier, and R. Ossenkoppele, "The effect of physical activity on cognitive function in patients with dementia: a meta-analysis of randomized control trials," *Ageing research reviews*, vol. 25, pp. 13-23, 2016.
- [14] E. Aguirre, R. T. Woods, A. Spector, and M. Orrell, "Cognitive stimulation for dementia: a systematic review of the evidence of effectiveness from randomised controlled trials," *Ageing research reviews*, vol. 12, pp. 253-262, 2013.
- [15] N. Farina, J. Rusted, and N. Tabet, "The effect of exercise interventions on cognitive outcome in Alzheimer's disease: a systematic review," *International Psychogeriatrics*, vol. 26, pp. 9-18, 2014.
- [16] J. Cohen-Mansfield, M. S. Marx, M. Dakheel-Ali, and K. Thein, "The use and utility of specific nonpharmacological interventions for behavioral symptoms in dementia: an exploratory study," *The American Journal of Geriatric Psychiatry*, vol. 23, pp. 160-170, 2015.
- [17] C. Anderson-Hanley, P. J. Arciero, A. M. Brickman, J. P. Nimon, N. Okuma, S. C. Westen, M. E. Merz, B. D. Pence, J. A. Woods, and A. F. Kramer, "Exergaming and older adult cognition: a cluster randomized clinical trial," *American journal of preventive medicine*, vol. 42, pp. 109-119, 2012.

- [18] T. Shibata and K. Wada, "Robot therapy: a new approach for mental healthcare of the elderly—a mini-review," *Gerontology*, vol. 57, pp. 378-386, 2011.
- [19] S. Coradeschi, A. Cesta, G. Cortellessa, L. Coraci, J. Gonzalez, L. Karlsson, F. Furfari, A. Loutfi, A. Orlandini, and F. Palumbo, "Giraffplus: Combining social interaction and long term monitoring for promoting independent living," in *Human system interaction (HSI), 2013 the 6th international conference on*, 2013, pp. 578-585.
- [20] D. McColl, W. Y. G. Louie, and G. Nejat, "Brian 2.1: A socially assistive robot for the elderly and cognitively impaired," *Robotics & Automation Magazine, IEEE*, vol. 20, pp. 74-83, 2013.
- [21] R. Looije, M. A. Neerinx, and F. Cnossen, "Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors," *International Journal of Human-Computer Studies*, vol. 68, pp. 386-397, 2010.
- [22] J. Le Maitre and M. Chetouani, "Self-talk discrimination in human–robot interaction situations for supporting social awareness," *International Journal of Social Robotics*, vol. 5, pp. 277-289, 2013.
- [23] Y. Matsusaka, H. Fujii, T. Okano, and I. Hara, "Health exercise demonstration robot TAIZO and effects of using voice command in robot-human collaborative demonstration," in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, 2009, pp. 472-477.
- [24] I. Bäck, K. Makela, and J. Kallio, "Robot-Guided Exercise Program for the Rehabilitation of Older Nursing Home Residents," *Annals of Long-Term Care: Clinical Care and Aging*, vol. 21, pp. 38-41, 2013.
- [25] T. Klamer and S. B. Allouch, "Acceptance and use of a social robot by elderly users in a domestic environment," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2010 4th International Conference on-NO PERMISSIONS*, 2010, pp. 1-8.
- [26] T. Shibano, Y. Ho, Y. Kono, Y. Fujimoto, and T. Yamaguchi, "Daily support system for care prevention by using interaction monitoring robot," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, 2010, pp. 3477-3482.
- [27] M. Kanoh, Y. Oida, Y. Nomura, A. Araki, Y. Konagaya, K. Ihara, T. Shimizu, and K. Kimura, "Examination of practicability of communication robot-assisted activity program for elderly people," *Journal of Robotics and Mechatronics*, vol. 23, p. 3, 2011.
- [28] B. Görer, A. A. Salah, and H. L. Akın, "A robotic fitness coach for the elderly," in *International Joint Conference on Ambient Intelligence*, 2013, pp. 124-139.
- [29] M. Simonov and G. Delconte, "Humanoid assessing rehabilitative exercises," *Methods of information in medicine*, vol. 54, pp. 114-121, 2015.
- [30] P. Gadde, H. Kharrazi, H. Patel, and K. F. MacDorman, "Toward monitoring and increasing exercise adherence in older adults by robotic intervention: a proof of concept study," *Journal of Robotics*, 2011.
- [31] K. Yoshino, M. Kouda, and Z. Shanjun, "Correct Motion Advice on Rehabilitation Instruction Robot by Superimposing Instructor CG Model," in *Intelligent Networks and Intelligent Systems (ICINIS), 2012 Fifth International Conference on*, 2012, pp. 333-336.
- [32] D. McColl and G. Nejat, "Meal-time with a socially assistive robot and older adults at a long-term care facility," *Journal of Human-Robot Interaction*, vol. 2, pp. 152-171, 2013.
- [33] A. Tapus, C. Tapus, and M. Matarić, "Long Term Learning and Online Robot Behavior Adaptation for Individuals with Physical and Cognitive Impairments," in *Field and Service Robotics*, 2010, pp. 389-398.
- [34] W. Y. G. Louie, T. Vaquero, G. Nejat, and J. C. Beck, "An autonomous assistive robot for planning, scheduling and facilitating multi-user activities," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, 2014, pp. 5292-5298.
- [35] J. Fasola and M. J. Matarić, "A Socially Assistive Robot Exercise Coach for the Elderly," *Journal of Human-Robot Interaction*, vol. 2, pp. 3-32, 2013.

- [36] R. Khosla and M.-T. Chu, "Embodying care in Matilda: an affective communication robot for emotional wellbeing of older people in Australian residential care facilities," *ACM Transactions on Management Information Systems (TMIS)*, vol. 4, p. 18, 2013.
- [37] Y. Matsuyama, H. Taniyama, S. Fujie, and T. Kobayashi, "Framework of Communication Activation Robot Participating in Multiparty Conversation," in *AAAI Fall Symposium: Dialog with Robots*, 2010.
- [38] R. Bemelmans, G. J. Gelderblom, P. Jonker, and L. De Witte, "Socially assistive robots in elderly care: A systematic review into effects and effectiveness," *Journal of the American Medical Directors Association*, vol. 13, pp. 114-120. e1, 2012.
- [39] R. Yu, E. Hui, J. Lee, D. Poon, A. Ng, K. Sit, K. Ip, F. Yeung, M. Wong, and T. Shibata, "Use of a Therapeutic, Socially Assistive Pet Robot (PARO) in Improving Mood and Stimulating Social Interaction and Communication for People With Dementia: Study Protocol for a Randomized Controlled Trial," *JMIR research protocols*, vol. 4, 2015.
- [40] W. Moyle, M. Cooke, E. Beattie, C. Jones, B. Klein, G. Cook, and C. Gray, "Exploring the effect of companion robots on emotional expression in older adults with dementia: A pilot randomized controlled trial," *Journal of gerontological nursing*, 2013.
- [41] H. Robinson, E. Broadbent, and B. MacDonald, "Group sessions with Paro in a nursing home: Structure, observations and interviews," *Australasian journal on ageing*, 2015.
- [42] N. Jøranson, I. Pedersen, A. M. M. Rokstad, and C. Ihlebæk, "Effects on symptoms of agitation and depression in persons with dementia participating in robot-assisted activity: a cluster-randomized controlled trial," *Journal of the American Medical Directors Association*, vol. 16, pp. 867-873, 2015.
- [43] H. Robinson, B. A. MacDonald, N. Kerse, and E. Broadbent, "Suitability of healthcare robots for a dementia unit and suggested improvements," *Journal of the American Medical Directors Association*, vol. 14, pp. 34-40, 2013.
- [44] K. Takayanagi, T. Kirita, and T. Shibata, "Comparison of verbal and emotional responses of elderly people with mild/moderate dementia and those with severe dementia in responses to seal robot, PARO," *Frontiers in aging neuroscience*, vol. 6, p. 257, 2014.
- [45] S. Petersen, S. Houston, H. Qin, C. Tague, and J. Studley, "The utilization of robotic pets in dementia care," *Journal of Alzheimer's Disease*, vol. 55, pp. 569-574, 2017.
- [46] J. Li, W.-Y. G. Louie, S. Mohamed, F. Despond, and G. Nejat, "A user-study with tangy the bingo facilitating robot and long-term care residents," in *2016 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS)*, 2016, pp. 109-115.
- [47] B. Görer, A. A. Salah, and H. L. Akın, "An autonomous robotic exercise tutor for elderly people," *Autonomous robots*, vol. 41, pp. 657-678, 2017.
- [48] J. Fan, L. Beuscher, P. A. Newhouse, L. C. Mion, and N. Sarkar, "A robotic coach architecture for multi-user human-robot interaction (RAMU) with the elderly and cognitively impaired," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, 2016, pp. 445-450.
- [49] J. Fan, D. Bian, Z. Zheng, L. Beuscher, P. A. Newhouse, L. C. Mion, and N. Sarkar, "A Robotic Coach Architecture for Elder Care (ROCARE) based on multi-user engagement models," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, pp. 1153-1163, 2017.
- [50] L. M. Beuscher, J. Fan, N. Sarkar, M. S. Dietrich, P. A. Newhouse, K. F. Miller, and L. C. Mion, "Socially Assistive Robots: Measuring Older Adults' Perceptions," *Journal of gerontological nursing*, vol. 43, pp. 35-43, 2017.
- [51] C. M. van Ravenswaaij-Arts, L. A. Kollee, J. C. Hopman, G. B. Stoeltinga, and H. P. van Geijn, "Heart rate variability," *Annals of internal medicine*, vol. 118, pp. 436-447, 1993.
- [52] J. T. Cacioppo, L. G. Tassinary, and G. Berntson, *Handbook of psychophysiology*: Cambridge University Press, 2007.
- [53] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm, "Looking at pictures: Affective, facial, visceral, and behavioral reactions," *Psychophysiology*, vol. 30, pp. 261-273, 1993.

- [54] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *Journal of biomedical informatics*, vol. 73, pp. 159-170, 2017.

## CHAPTER 6

### MATHEMATICAL MODEL OF ADAPTATION FOR MULTI-USER HRI

#### 6.1 Brief Summary

The models of people, interaction, and machine are building blocks for adaptive automation of SAR systems. In CHAPTER 2 and CHAPTER 3, I have designed and developed models of people and interaction. This chapter presents my research towards the design of a mathematical model for multi-user HRI that integrates a model of people, a model of interaction, and a model of machine.

This model is specifically designed for SAR applications with an emphasis on enforcing human-human interaction (HHI) through HRI. We first defined formally a set of multi-user HRI scenarios with well-defined task, goal, and robot role, which we referred to as problem formulation. In this problem formulation, we modelled human behaviors as interaction modes which were represented by human attentiveness (HA) and human cooperative level (HCL). We then further represented HA and HCL by human intention (HI), which is context-sensitive and could interpret both mental state and task-related information. In addition, we added another human property, human adaptability to the robot, to serve as the human learning model. With the definition of the task, the robot role, and the interaction mode represented by HA, HI, HCL, and adaptability as the essential elements in multi-user HRI, we formalized the multi-user HRI as a mixed-observable Markov decision process (MOMDP).

To conduct model simulation, we modified a *RockSample* problem as our multi-user HRI scenario. A user simulator was designed to mimic the strategic behaviors of humans in the *RockSample* game. The user simulator generates the play actions for users conditioned on each user's level of cooperation with other users and with the robot, and each user's noise level. The simulator was implemented using reinforcement learning and did not know how the MOMDP model of the *RockSample* game performs online planning for the next robot action. Simulation results demonstrated the ability of the model to estimate and shape HA, HI, and HCL, learn human adaptability, and facilitate task completion.

#### 6.2 Abstract

There are different forms of human-robot interaction (HRI), which are one robot interacts with one human, multiple robots interact with multiple humans, multiple robots interact with one human, and one robot interacts with multiple humans. In this paper, we focused on the application of HRI in the forms of

socially assistive robotic (SAR) systems with an emphasis on enforcing human-human interaction (HHI) through HRI. Within this context, we propose to formally model multi-user HRI with well-defined task formulation and a mathematical model of people, interaction, robot, and their integration. We believe this is the first work that formalize the interaction flow among human users and how robot actions could shape that interaction flow. This mathematical model is designed for the control of SAR that takes into account individual differences and is applicable for reuse across a variety of interaction scenarios that confirm to our problem formulation.

### 6.3 Introduction

Human-robot interaction (HRI) problems are not restricted to one-on-one interaction. Many researchers are developing systems that allow for dynamic interactions in human-robot teams. The main applications of human-robot teams are in the domains such as search and rescue, and unmanned/uninhabited air vehicles (UAVs). Research in these domains has focused on multi-robot cooperative localization, object tracking, target detection, and perception [1-4], human-robot trust [5-7], situation awareness [6, 8], user interface design [6, 9], and operator mental workload [4, 7]. The structure of the human-robot teams in these works is usually one or few human manage(s) multiple robots, and the role of the robots are often peers or tools. Socially assistive robotic (SAR) systems, on the contrary, are often designed to become a mentor or companion around human. Human-robot teams in this case are likely to have the structure of a group of human interacting with one or few robots. Kanda et al. [10] designed Robovie, an interactive robotic classroom tutor, and tested it in an elementary school. A health exercise robot TAIZO was developed by Matsusaka et al. to demonstrate physical exercises [11]. Another application of multi-user SAR systems is robotic tour guides [12, 13]. Louie et al. [14] developed robot Tangy to schedule and play Bingo games with multiple users. Keizer et al. [15] developed a bartender robot to serve drinks to multiple users. Kondo et al. [16] developed an interaction system on an android Actroid-SIT for communication with multiple people. Although multiple human were involved in HRI, these SAR systems were not designed with an intention to observe or provide feedbacks based on the interaction flow among human or the group behavior.

There is a growing interest in multi-user HRI with an emphasis on interaction flow among human and group behavior. Chandra et al. [17] used a learning-by-teaching paradigm with a NAO robot facilitator (Wizard of Oz setting) to support the interaction flow between two children. Alves-Oliveira et al. [18] analyzed the dialogue utterances generated during a collaborative learning task played by two students and a teacher in order to develop dialogue dimensions for a future robotic tutor. Another work by Alves-Oliveira et al. [19] studied the emotions of a group for the purpose of an emotionally intelligent robotic agent for

multi-user HRI. Similarly, Leite et al. compared models of disengagement in individual and group interactions [20] using data collected with their multi-robot single-user/multi-user system [21]. Yumak et al. [22] developed a multi-party interactive system consisted of a virtual character, a humanoid robot, and two human users. The interaction flow between two users, users and virtual character/robot was addressed by multi-user tracking and fusion. Matsuyama et al. [23] developed a SAR system with robot Schema to promote the communication activeness among a group of people. The robotic classroom tutor Robovie [24] estimated friendly relationships among people through observation of group behavior. Vazquez et al. [25] proposed a robot-centric state representation and explored reinforcement learning methods to control the attention of a robot, through robot's orientation, during simulated multi-party conversations.

In addition to a lack of literature on multi-user HRI, the existing works have focused on describing and predicting responses of human to HRI through observation or sensor fusion, forming the hierarchy of the interaction, and are usually task-specific<sup>3</sup>. We believe that formalized models are needed to understand the goals, tasks, and interaction flow of the multi-user HRI, and shape the interaction flow within the human-robot teams. To the best of our knowledge, there are no formalized models proposed in the literature for multi-user HRI. In our previous work, we proposed a multi-user robotic coach architecture for elder care, ROCARE, which was based on the sense-think-act paradigm with mathematical definitions of each module, relationships among modules, and engagement models that capture the dynamics of interaction between human users, and between human and robot [26]. Together with a formalized model, the goals, tasks, and interaction flow during multi-user HRI could be quantified for better and more robust system performance. In this work, we present a formalized model for multi-user HRI where one robot interacts with a group of people.

There are existing works on formalized models in HRI and human factors. Nikolaidis et al. [27] proposed a generalized human-robot mutual adaptation formalism. The mutual adaptation was formulated in shared-autonomy setting as a Mixed Observability Markov Decision Process (MOMDP) model. The robot goal was to guide the human operator towards the optimal task goal and retain human trust at the same time. The human goal was modeled as a latent variable that changed values based on human belief on the robot goal, which was estimated by the robot, and human adaptability. This model allowed the robot to infer both the human goal and the human adaptability. If the human adaptability was high, the robot guided the human towards an optimal task. Otherwise, the robot followed the suboptimal human goal to retain human trust. Instead of assuming full adaptation of human, Nikolaidis et al. [28] presented a model of human partial adaptation in human-robot collaboration setting. In this model, human did not have the

---

<sup>3</sup> One exception is Vazquez et al. [25], who exploited the spatial organizations of group conversations and explored reinforcement learning methods to control robot's orientation in a simulated environment.



knowledge of robot's true capabilities, which was represented as a reward matrix. Through repeated play of the same game, human gradually learned rows of reward matrix that robot had played with a learning rate. This model allowed the robot to decide whether to reveal information to human or play the best action given the current human understanding of the reward matrix. Leveraging the concept of intentional reasoning and theory of mind, Baker et al. modeled the human action understanding [29], human plan recognition [30], and human social goal inference [31] when observing one or multiple agents moving in some environments. The theory-based models could infer an agent's belief and desire, or one agent's goal adopted in relation to another agent based on the actions taken in the environment by the agent(s). Building on top of [31], Ullman et al. [32] proposed a framework for modeling more complex social goals such as one agent helping or hindering another agent. An active model was proposed by Sadigh et al. [33] where the system took active information gathering actions to estimate the implicit human internal state. The robot inferred human internal state based on the effects of its actions on the human actions. Wang [4] presented a decision-making framework for the allocation of autonomous agents sensing mode and manual human sensing mode in search tasks using a human-agent collaborative team. In this framework, a human workload and sensing capability model was proposed to capture the relationship among human performance, task difficulty, and human utilization ratio.

Our multi-user HRI model is specifically designed for SAR applications with an emphasis on enforcing human-human interaction (HHI) through HRI. We propose to formally model multi-user HRI with well-defined task formulation and a mathematical model of people, interaction, robot, and their integration. The first step is to formally define a set of tasks, which we refer to as problem formulation. This is the key step towards a formalized model. Without a well-defined set of tasks, it is impossible to design a mathematical model. Based on the problem formulation, we then formally model the people, the interaction, and the robot for multi-user HRI. The mathematical model for one-on-one HRI is a special case of multi-user HRI. The main contribution of our model is to formalize the interaction flow among human users and how robot actions could shape that interaction flow. In the one-on-one HRI case [27, 28, 33], the only interaction flow is between the human and the robot. Although the work by Baker et al. [31] and Ullman et al. [32] incorporated multiple agents and their social goals, i.e., one agent's goal depended on another agent, these social goals were static throughout the interaction. In contrast, user states or interaction modes in our model are dependent on each other's actions and the robot's actions. The goal of the robot is to reason over the interaction modes of the users and take actions to shape the interaction modes. We further contribute by presenting a human learning model that is generalized from [28] to multi-user HRI. The current model is a passive approach, in the future we would like to explore how to incorporate active information gathering in our model.

## 6.4 Computational Framework

### 6.4.1 Problem formulation

Consider a multi-user HRI scenario where a group of human users perform a task collaboratively with the help of a robot. The goal of the robot is to support the interaction flow between the users through HRI in the context of a collaborative task. Collaborative tasks that require communication between human users are useful for fostering socialization among isolated individuals such as older adults with apathy or promoting the development of social skills of children. We use two users-one robot interaction as an example to demonstrate the interaction flows in a task. The problem formulation is generalized to any number of users. Figure 6-1 illustrates several possible configurations of the interaction flows among the human-robot team and the world. Each human user may engage in the task and thus establish interaction flow between the user (H) and the world (W). If two users perform a task collaboratively, there is an interaction flow between the users. As the task requires HHI, the robot establishes interaction flows between itself (R) and H, and between R and W to facilitate task completion. The dynamics of the interaction flow is determined by user states of each human user, which we refer to as interaction modes, and is shaped by the robot and the user actions through HRI and HHI. In Figure 6-1, the preference for the configurations following the order:  $(a) > (b) > (c) > (d)$ . The role of the robot gradually fades away as it successfully engages human users in the task and HHI. We define two human interaction modes related to the configuration of the interaction flows, which are i) user engagement with respect to the task, or human attentiveness (HA), and ii) user engagement with respect to collaboration in task, or human cooperative level (HCL). If HA is high, there is an effective interaction flow between H and W. If HCL is high for the two users, there is an effective interaction flow between H and H.

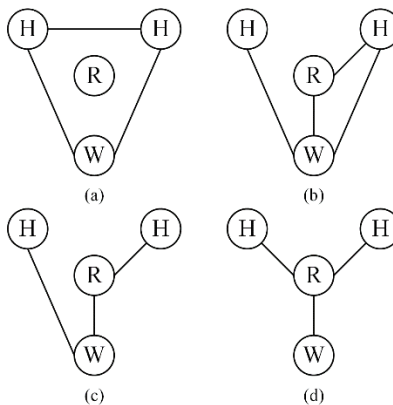


Figure 6-1. Examples of Interaction Flows among the Human-Robot Team

Based on the abovementioned problem setting, we formulate the problem as follows:

### *Task*

- i) Users have and are aware of common start and end task states.
- ii) There is a discrete and countable number of intermediate goals associated with the task. In the process of completing the task, at each time step, human users may have a singular and unambiguous intermediate goal. We refer to their tendency to fulfil their intermediate goal as intention.
- iii) The task has components that create gaps in users' intentions to enforce HHI. In other words, users' understandings on how to get to the end task state from the start task state by choosing immediate goals may not match at the beginning. For example, one user knows the route from A to B whereas the other user may not. A more realistic example is that two users are considering different routes from A to B, however, they need to share a vehicle. We refer to each user's intentions as human intention (HI). HHI, such as negotiation, or robot actions could shrink the gap between users' HIs. With matched HIs, users can collaborate efficiently and thus have high successful rate in task completion.

### *Goal*

- i) The end task state is the common goal shared by all the users. This common goal is the eventual reward that all the users agree with.
- ii) Based on each user's HI, they have different intermediate goals, or some users do not have intermediate goals.

### *Robot Role*

- i) The robot needs to estimate each user's HI, HA, HCL, and human adaptability based on interaction history, and shape HI, HA, and HCL through HRI.
- ii) The robot needs to serve as an arbitrator to help shrink the gap between users' HIs.
- iii) The robot needs to help with task completion in the event one or multiple users are not contributing to the task completion.

This problem formulation represents humans using three variables, which are HA, HI, and HCL. The unconstrained human behaviors during multi-user HRI are represented by interaction modes HA and HCL, which can be represented by HI. Thus, HI is first-order representation whereas HA and HCL are second-order representation. Together, they are used to model human states and interactions during multi-user HRI.

#### 6.4.2 Formal modeling

The mathematical model is structured to be a mixed-observability Markov decision process (MOMDP) [34] as a tuple  $\{X, Y, A, T_x, T_y, R, \Omega, O\}$ :

$X : X_{world} \times \{A_i\}_{i=1}^n$  is the observable variable. It represents the current world state or task state. Human actions at the current time step are stored in the observable variable. If the desired action for the robot is to observe human interacting with the world, then human actions are used to determine the transition of the world state.  $n$  denotes the number of users.

$Y : \{B_i \times HA_i \times HI_i\}_{i=1}^n$  is the set of partially observable variables. These are the HA and HI for each human user. HA is a Boolean variable that are either true or false. When HA is false, the value of HI does not matter, otherwise HI is one of the possible intentions.  $B_i$  represents each user's adaptability to robot actions to influence one's HA and HI.

$A : \{A_{r \rightarrow x}, A_{r \rightarrow y}\}$  is a finite set of actions. These are the robot actions that relate to the transitions of the observable variable  $X$ , represented by  $A_{r \rightarrow x}$ , and that relate to transitions of the set of partially observable variables  $Y$ , represented by  $A_{r \rightarrow y}$ . We assume that the robot can only perform one action at a time, either change world state or change human states.

$T_x : X \times A_{r \rightarrow x} \rightarrow X$  is a deterministic mapping from a previous task state  $x_{world}$ , collective human actions  $\{a_i\}_{i=1}^n$  and robot action  $a_{r \rightarrow x}$ , to a subsequent observable state  $x'$ . We assume when robot takes actions to change  $X$ , human actions are suppressed. Otherwise, the combined human actions change  $X$ .

$T_y : Y \times A_{r \rightarrow y} \rightarrow \Pi(Y)$  is the probability of the human switching to a different human attentiveness  $ha$  or human intention  $hi$  given robot action  $a_{r \rightarrow y}$  and human adaptability  $\{\beta_i\}_{i=1}^n$ .

$R: X \times \{HA_i \times HI_i\}_{i=1}^n \times A \rightarrow \mathbb{R}$  is a reward function that gives an immediate reward for human attentiveness  $ha$ , human cooperative level  $hcl$  as represented by the similarity between users' intentions  $Similarity: \{HI_i\}_{i=1}^n \rightarrow HCL$ , task state  $x_{world}$ , and robot actions  $a_{r \rightarrow x}$  and  $a_{r \rightarrow y}$ .

$\Omega: \{A_i\}_{i=1}^n$  is the set of observations that the robot perceives about what actions human users' take.

$O: \{HA_i \times HI_i\}_{i=1}^n \times X \rightarrow \prod(\Omega)$  is the observation function, which gives a probability distribution over human actions for a set of human attentiveness  $ha$  and human intentions  $hi$  as well as the current task state  $x_{world}$ .

The variables and functions  $X, Y, A, T_x, \Omega$  can be defined based on the task, in the following paragraphs, I propose the equations for functions  $T_y, O, R$ .

$T_y$  captures how robot actions could shape human attentiveness HA, human intention HI, as well as human cooperative level HCL. In this model, HCL is not included as one of the hidden variables. There are two reasons for this decision. First, in the problem formulation section, HCL is defined to have direct relationship with HI, therefore HCL can be represented by HI. Second, using HCL as a partially observable variable will dramatically increase the complexity of the model. As we increase the number of users, partially observable variables, HA and HI, increase linearly. If we define a HCL variable between every two users, the size of HCL variables becomes  $C_n^2$  and therefore HCL increases quadratically.  $T_y$  has two sets of equations, one for HA and one for HI. When HA is false for user  $i$  ( $ha_i = 0$ ), user  $i$ 's intention  $hi_i$  does not matter and therefore there is no collaboration between user  $i$  and any other users. Let us define robot action to increase  $ha_i$  as  $a_{r \rightarrow ha_i}$ :

$$\Pr(ha_i' | ha_i, a_{r \rightarrow ha_i}) = \begin{cases} \beta_i & \text{if } ha_i' = 1 \text{ and } ha_i = 0 \\ 1 - \beta_i & \text{if } ha_i' = 0 \text{ and } ha_i = 0 \end{cases} \quad (6.1)$$

If HA is true for user  $i$  ( $ha_i = 1$ ), then robot actions to foster HHI become effective. We first define a similarity function that represents the similarity between users' intentions. HCL between users is then represented as a function of similarities between users' intentions  $Similarity: \{HI_i\}_{i=1}^n \rightarrow HCL$ . If there are two users, then HCL is equal to  $Similarity(HI_1, HI_2)$ . If  $n > 2$ , depending on the HCL definition for a

particular many-to-one HRI scenario, HCL could be defined by different combinations of the similarity functions. For example, HCL could take the value  $\sum_i^n \sum_j^n w_{ij} \text{Similarity}(HI_i, HI_j)$  that evaluates HI between each pair of users. Another possible HCL function could evaluate the similarities of HIs among a subset of users. With the definition of HCL, we can then define  $T_y$  with  $ha_i = 1$  as:

$$\Pr\left(hi'_1, \dots, hi'_n \mid hi_1, \dots, hi_n, a_{r \rightarrow hi}\right) = \begin{cases} \beta & \text{if } hcl_n \neq hcl_{\max} \wedge \left( hcl'_n - hcl_n > \text{margin} \vee hcl'_n = hcl_{\max} \right) \\ 1 - \beta & \text{if } hcl_n \neq hcl_{\max} \wedge \left( hcl'_n - hcl_n \leq \text{margin} \wedge hcl'_n \neq hcl_{\max} \right) \\ 1 & \text{if } hcl_n = hcl_{\max} \wedge hcl'_n = hcl_{\max} \\ 0 & \text{if } hcl_n = hcl_{\max} \wedge hcl'_n \neq hcl_{\max} \end{cases} \quad (6.2)$$

Note that unlike HA, HCL can be a continuous variable if the similarity function outputs a continuous variable. The value of HCL will be bounded to be in the range  $[hcl_{\min}, hcl_{\max}]$ . Equation 6.2 presents the  $T_y$  when the task defines HCL as the collaboration of the whole group. In this case,  $\beta$  is the combination of  $\{\beta_i\}_{i=1}^n$ . The  $T_y$  can be further expanded to incorporate HCL among subgroups. The margin in Equation 6.2 is a hyper-parameter that guards the amount of changes of HCL we consider as effective given a robot action.

Next we define the observation function  $O$ . Similar to  $T_y$ ,  $O$  has two sets of equations, one for  $ha_i = 0$  and one for  $ha_i = 1$ . When  $ha_i = 0$ , user  $i$  is not paying attention to the task, and therefore user  $i$  takes actions randomly.

$$\Pr(a_i \mid hi_i, ha_i = 0, x_{\text{world}}) = 1/m \quad (6.3)$$

where  $m$  is the number of possible human actions. When  $ha_i = 1$ , user  $i$  takes actions according to one's  $hi_i$  and the current task state  $x_{\text{world}}$ .  $\Pr(a_i \mid hi_i, ha_i = 1, x_{\text{world}})$  is computed by formulating the task as a Markov decision process (MDP) with  $hi_i$  as the end goal, and solving the probability of a human action by applying a softened version of MDP which incorporates the principle of maximum entropy [35, 36]. Assume each human user takes actions according to their  $ha_i$  and  $hi_i$  and independent from actions of other

users, the observation function  $O$  becomes  $\Pr\left(\{a_i\}_{i=1}^n \mid \{hi_i\}_{i=1}^n, \{ha_i\}_{i=1}^n, x_{\text{world}}\right) = \sum_{i=1}^n \Pr(a_i \mid hi_i, ha_i, x_{\text{world}})$ .

Finally, we define the reward function  $R$  to have four components:

$$R(x, \{ha_i, hi_i\}_{i=1}^n, a_{r \rightarrow x}, a_{r \rightarrow y}) = R(\{ha_i\}_{i=1}^n)R(hcl_n)R(task)R(a_r) \quad (6.4)$$

The first component  $R(\{ha_i\}_{i=1}^n)$  represents the need to make sure human users are engaged with respect to the task. The second component  $R(hcl_n)$  makes sure the robot takes actions to foster HHI. The third component is related to the reward for making progress to complete the task. The last component is the cost associated with each robot action. For one-to-one HRI, the model can be adapted by removing the similarity function, the transition function for HI, and the reward associated with HCL.

## 6.5 Simulation

### 6.5.1 Multi-user HRI scenario

We modified the *RockSample* problem [37] as our concrete HRI scenario. Figure 6-2 illustrates an example of the *RockSample* task environment. The task is to move a rover agent in the grid world and collect all the rock samples. The positions of the agent and the rocks are known to all the users. Each user can control the agent with five actions, which are  $\{east, west, north, south, stay\}$ . The combined user actions determines where users want to move the agent next. The final action the agent takes depends also on a robot's decision, who facilitates the users to accomplish this cooperative task. If the robot decides to follow human commands, then the agent in the task moves according to the combined user actions. If the robot decides to move the agent by itself, then the users' actions are ignored. If the robot decides to provide feedback to the users, then the agent does not move. A rock is collected if the agent moves to the position of the rock. The task ends when all the rocks are collected by the agent.

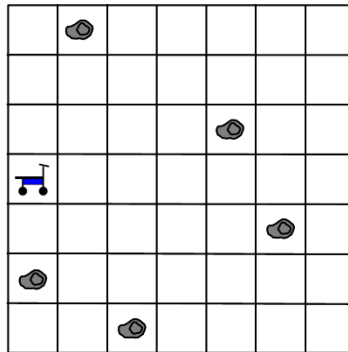


Figure 6-2. Example *RockSample* Task Environment

The task complies with the problem formulation and has the following properties:

- i) Users have and are aware of common start and end task states. For example, users know the start position of the agent and the end state is reached when all the rocks are collected.
- ii) Users have clear intentions or intermediate goals. Users need to select which rock they want to collect next.
- iii) Users are likely to have different intentions. This is achieved by the fact that each user might prefer to collect a different rock and to collect them in different orders.

### 6.5.2 The MOMDP model of *RockSample*

An instance of *RockSample* with map size  $M \times M$ , user size  $N$ , and  $K$  rocks is described as  $RockSample[M, N, K]$ . The MOMDP models of  $RockSample[M, N, K]$  is as follows. The state space is the observable state variables and partially observable state variables. The observable states is the cross product of  $K + 2$  features:  $Position = \{(1,1), (1,2), \dots, (m,m)\}$ , combined human actions to follow  $A_h = \{east, west, north, south, stay\}$ , and  $K$  binary features  $RockExist_j = \{True, False\}$  that indicate which rock has not been collected. The partially observable states is the cross product of  $3N$  features: adaptability  $B_i \in \{0, 0.25, 0.50, 0.75, 1.00\}$ , human attentiveness  $HA_i \in \{0, 1\}$ , and human intention  $HI_i \in \{1, \dots, K\}$  for each user. The terminal state is reached when  $\forall j \in K, RockExist = False$ .

The robot can select from  $6 + N$  actions:  $\{east, west, north, south\}$  when the robot controls the agent in the game by itself,  $\{follow\_humans\}$  when the robot controls the agent based on combined human actions  $A_h$ ,  $\{A_{r \rightarrow ha_1}, \dots, A_{r \rightarrow ha_N}\}$  when the robot tries to influence each user's  $ha$ , and  $\{A_{r \rightarrow hcl}\}$  when the robot tries to influence users'  $hcl$  by asking all the users to collect the rock that is closest, based on Euclidean distance, to the agent. The first 5 are deterministic single-step motion actions. The robot receives a reward of -50 if it ignores human actions and moves the agent by itself, and an additional reward of -100 if users' have high  $ha$  and  $hcl$  based on robot's own belief of the partially observable variables. The action  $A_{r \rightarrow ha_i}$  tries to influence one user's  $ha$ , the robot receives a reward of -10. If user  $i$  is already engaged based on the partially observable variable estimation, the robot receives another reward of -100. The robot updates its belief on  $ha_i$  based on  $\beta_i$ . If  $ha_i$  switches from 0 to 1, the robot receives a reward of 60. Finally, when the robot tries to influence  $hcl$ , it provides feedback to request all the users to switch



their  $hi_i$  to a new  $hi_{i'}$ , a rock that is closest to the agent. Initially, the robot receives a reward of -10. If any user is not engaged or  $hcl$  estimation is high at this time step, the robot receives another reward of -100. The robot updates its belief on  $hi_i$  based on  $\beta_i$  only if  $hi_i \neq hi_{i'}$ . If user switches their  $hi_i$  successfully, the robot receives a reward of 10. In terms of the *RockSample* game, the robot receives a reward of -100 each time the robot action moves the agent into a position that is outside of the grid world. The robot receives a reward of 10 if a rock is collected and when the game terminates.

At each time step, each human user provides their desired action for the agent to move in the game environment. If their actions are identical, then the combined action follows their desired actions. Otherwise, the combined action is  $A_h = stay$ . The observation variables are the observed human actions. The observation probability is implemented following model description in Section 6.4.2. The similarity function that computes  $hcl$  is implemented as cosine similarity and bounded to  $[-1,1]$ ,

$$hcl: \forall i \neq j, \binom{N}{2} \sum_{i=1}^N \sum_{j=1}^N \cos\_similarity(hi_i, hi_j), \text{ where } \cos\_similarity(hi_i, hi_j) = \frac{\vec{v}_{p_{hi_i}}^{Pr} \bullet \vec{v}_{p_{hi_j}}^{Pr}}{\|\vec{v}_{p_{hi_i}}^{Pr}\| \|\vec{v}_{p_{hi_j}}^{Pr}\|}, \vec{v}_{p_{hi_i}}^{Pr}$$

denotes the vector from current agent position  $p_t$  to the rock position  $p_{hi_i}$  that  $hi_i$  represents.

The DESPOT algorithm is used as a solver [38]. DESPOT is an online POMDP algorithm that performs heuristic search in a sparse belief tree conditioned under a set of sampled ‘‘scenarios’’. Each scenario comprises a sampled starting state and a stream of random numbers to determine future transitions and observations. For our multi-user HRI scenario, the observable states are deterministic. We initialize  $A_h$  to be *stay* and  $\forall j \in K, RockExist = True$ . For initialization of partially observable variables,  $B_i$  has a

$$\text{discrete probability distribution of } \Pr(\beta_i) = \begin{cases} 0.1 & \beta_i = 0 \\ 0.1 & \beta_i = 0.25 \\ 0.2 & \beta_i = 0.5 \\ 0.3 & \beta_i = 0.75 \\ 0.3 & \beta_i = 1.0 \end{cases}, HA_i \text{ has a discrete probability distribution of}$$

$$\Pr(ha_i) = \begin{cases} 0.25 & ha_i = 0 \\ 0.75 & ha_i = 1 \end{cases}, \text{ and } HI_i \text{ has a discrete probability distribution based on the distance of the agent}$$

to each rock,  $\Pr(hi_i^{(j)}) \propto \exp(-dist(p_t, p_{rock}^{(j)}))$ , where  $j$  denotes the  $j^{\text{th}}$  rock.

### 6.5.3 User simulation

To evaluate our work, a simulator is presented in this section, which is the external system interacting with our MOMDP model and mimicking the strategic behaviors of humans in the *RockSample* game.

The simulator includes a collection of players  $\{P_i\}_{i=1}^N$  (assume it has totally  $N$  players and the rock game has  $K$  unpicked rocks). Each player  $P_i$  is described by 4-tuple  $(h_i, r_i, n_i, t_i)$ , where

- $h_i \in [0,1]$  is the level of cooperation between player  $i$  and the other players. Higher  $h_i$  leads player  $i$  to be more inclined to follow the social choice. When  $h_i = 0$ , player  $i$  will ignore the behaviors of others completely. When  $\forall i = 1, \dots, N$ ,  $h_i = 1$ , all players will reach to a consensus in one game step.
- $r_i \in [0,1]$  is the level of cooperation between player  $i$  and the robot. Higher  $r_i$  means player  $i$  trend to more actively respond to the actions of the robot. When  $r_i = 0$ , player  $i$  will ignore the actions of the robot completely.
- $n_{i,t} \in \mathbb{R}^+$  is the noise level of player  $i$  at time step  $t$ . Player  $i$  with higher  $n_{i,t}$  will trend to make a random decision rather than a rational one and  $n_{i,t} = 0$  means player  $i$  is completely rational at that moment.
- $t_{i,t} \in [0,1]^K$  is the target belief of player  $i$  at time step  $t$ , which is a vector with  $K$  elements. Let  $t_{i,t}^{(j)}$  be the  $j^{\text{th}}$  element in  $t_{i,t}$ . Higher  $t_{i,t}^{(j)}$  indicates player  $i$  has a stronger intention to move to the position of the  $j^{\text{th}}$  rock. If player  $i$  has no interest to the  $j^{\text{th}}$  rock at the moment,  $t_{i,t}^{(j)} = 0$ .

Now we describe the simulation process. For each game iteration, players first submit their actions to the system. Let  $A_{i,t} \subset \{east, west, north, south, stay\}$  be the action space of player  $i$  at time step  $t$ ,  $p_t$  be the current position of the agent,  $\{p_{rock}^{(j)}\}_{j=1}^K$  be the positions of the rocks, and  $\forall a \in A_{i,t}, p_{a,t}$  be the positions after the agent applies the action  $a$ . Player  $i$  calculates the values of its possible actions according to the sum of the kernel distances from  $p_{a,t}$  to the positions of rocks plus the decision noise. Formally

$$V_{i,t}(a)_{a \in A_{i,t}} = \sum_{j=1}^K \exp\left(-dist\left(p_{rock}^{(j)}, p_{a,t}\right)\right) t_{i,t}^{(j)} + \mu_{i,t} \quad (6.5)$$

where  $dist$  is the  $l_1$  distance function (which equals to zero when the corresponding rock is picked) and  $\mu_{i,t} \sim N(0, n_{i,t})$  is a Gaussian random variable with variance  $n_{i,t}$ . Then, the action with maximal value will be the final decision of player  $i$ .

$$a_{i,t} = \arg \max_{a \in A_{i,t}} V_{i,t}(a) \quad (6.6)$$

Next, players update their properties according to the robot's action.

- In the cases where the robot ignores players' actions, the game will be played by the robot, players' actions will be discarded and their properties will keep unchanged.
- In the cases where the robot follows players' actions, if all of them make the same decision, the game will be played by players and players' properties will keep unchanged. Otherwise, the agent will not move and players will trend to change their target beliefs to the average (e.g. players may have a discussion and try to reach to a consensus). Formally,

$$\forall i, t_{i,t+1} = t_{i,t} + h_i (\bar{t}_i - t_{i,t}) \quad (6.7)$$

where  $\bar{t}_i = \frac{1}{N} \sum_{i=1}^N t_{i,t}$  is the average target belief among players.

- In the cases where the robot hints players to focus on the nearest rock, the agent will not move and players will trend to update their target beliefs according to the robot's hint. Formally,

$$\forall i, t_{i,t+1}^{(j)} = t_{i,t}^{(j)} + r_i \quad (6.8)$$

where the  $j^{\text{th}}$  rock is the nearest one to the agent and  $t_{i,t+1}$  will be normalized after each iterations, to make sure all its elements are between 0 and 1.

- In the cases where the robot engages player  $i$  to focus on the game, the agent will not move and player  $i$  will trend to reduce its level of noise. Formally,

$$n_{i,t+1} = (1 - \alpha r_i) n_{i,t} \quad (6.9)$$

where  $\alpha \in (0, 1]$  is a weight parameter.

### 6.5.4 Results

The model is able to run and adapt its beliefs on user’s HA, HI, HCL, and adaptability to the robot for two users. The simulator is able to simulate the play actions for two users conditioned on user’s level of cooperation with other users and with the robot, and user’s noise level. The model and the simulator are integrated together and the model performs online planning for the next robot action. We demonstrated the ability of the model to estimate and shape HA, HI, and HCL, learn human adaptability, and facilitate task completion with three simulation cases.

#### Case 1

In this case, we used task *RockSample*[4,2,4], the properties of the two players were defined as  $P_0 : (h = 0.99, r = 1.0, n = 50, t = [0, 0, 50, 0])$  and  $P_1 : (h = 0.99, r = 1.0, n = 0, t = [0, 0, 0, 50])$ . Both players had high cooperative level with each other and with the robot. The first player had high noise level. Their target belief on the rocks were different. The simulation results are shown in Figure 6-3. It can be seen that as humans take actions in the multi-user HRI task, the model gradually updates its belief on a player’s HA. By step 5, the model was certain that the first player had very low HA and thus chose action “Influence user 0 engagement” to provide feedback to improve the first player’s HA. This simulation demonstrated the model’s ability to estimate and shape HA. The whole task was completed in 12 steps and cost 24.65s.

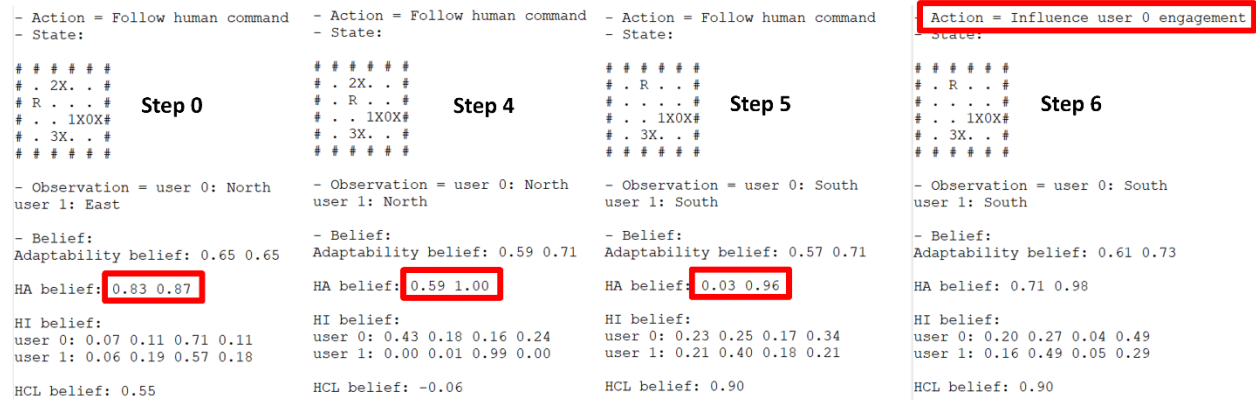


Figure 6-3. Case 1 Simulation Results

## Case 2

In this case, we used task *RockSample*[4,2,4], the properties of the two players were defined as  $P_0 : (h = 0.1, r = 1.0, n = 0, t = [0, 0, 50, 0])$  and  $P_1 : (h = 0.1, r = 1.0, n = 0, t = [0, 0, 0, 50])$ . Both players had low cooperative level with each other, however, they had high cooperative level with the robot. Both players were rational players and their target belief on the rocks were different. The simulation results are shown in Figure 6-4. It can be seen that as humans take actions in the multi-user HRI task, the model gradually updates its belief on their HA, HI, and HCL. By step 4, the model learned that both players had high HA, the first player's intermediate goal was to collect the third rock and the second player's intermediate goal was to collect the fourth rock. Therefore, the model chose action "Influence human collaboration" at step 5 to provide feedback to improve players' HCL by shrinking the gap of their HIs. This simulation demonstrated the model's ability to estimate and shape HI and HCL. The whole task was completed in 12 steps and cost 25.44s.

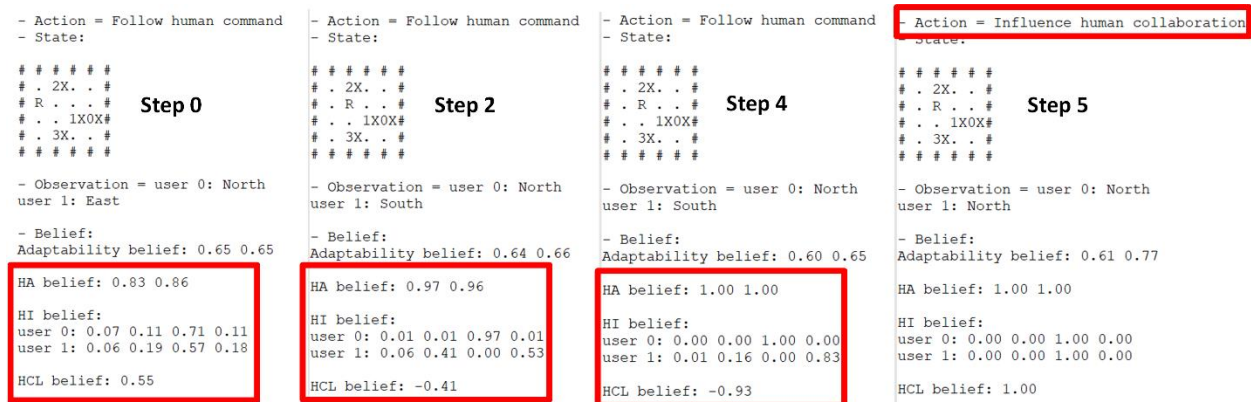


Figure 6-4. Case 2 Simulation Results

## Case 3

In this case, we used task *RockSample*[4,2,4], the properties of the two players were defined as  $P_0 : (h = 0.01, r = 0.2, n = 50, t = [0, 0, 50, 0])$  and  $P_1 : (h = 0.01, r = 0.2, n = 0, t = [0, 0, 0, 50])$ . Both players had low cooperative level with each other and with the robot. The first player had high noise level. Their target belief on the rocks were different. The simulation results are shown in Figure 6-5. It can be seen that as humans take actions in the multi-user HRI task, the model learned that the first player had low HA at step 3 and provided action to influence this player's HA accordingly at step 4. However, since the level of

cooperation between this player and the robot was low, the model realized that the first player’s HA did not improve much due to its action to improve engagement. As the model continue interacting with the user simulator and updating its belief on the two players, it learned that the first player had low adaptability to robot feedback (step 12 and step 14). Eventually, the model decided to take action “East” to facilitate task completion at step 21. This simulation demonstrated the model’s ability to learn human adaptability and to facilitate task completion. The whole task was completed in 25 steps and cost 47.82s.

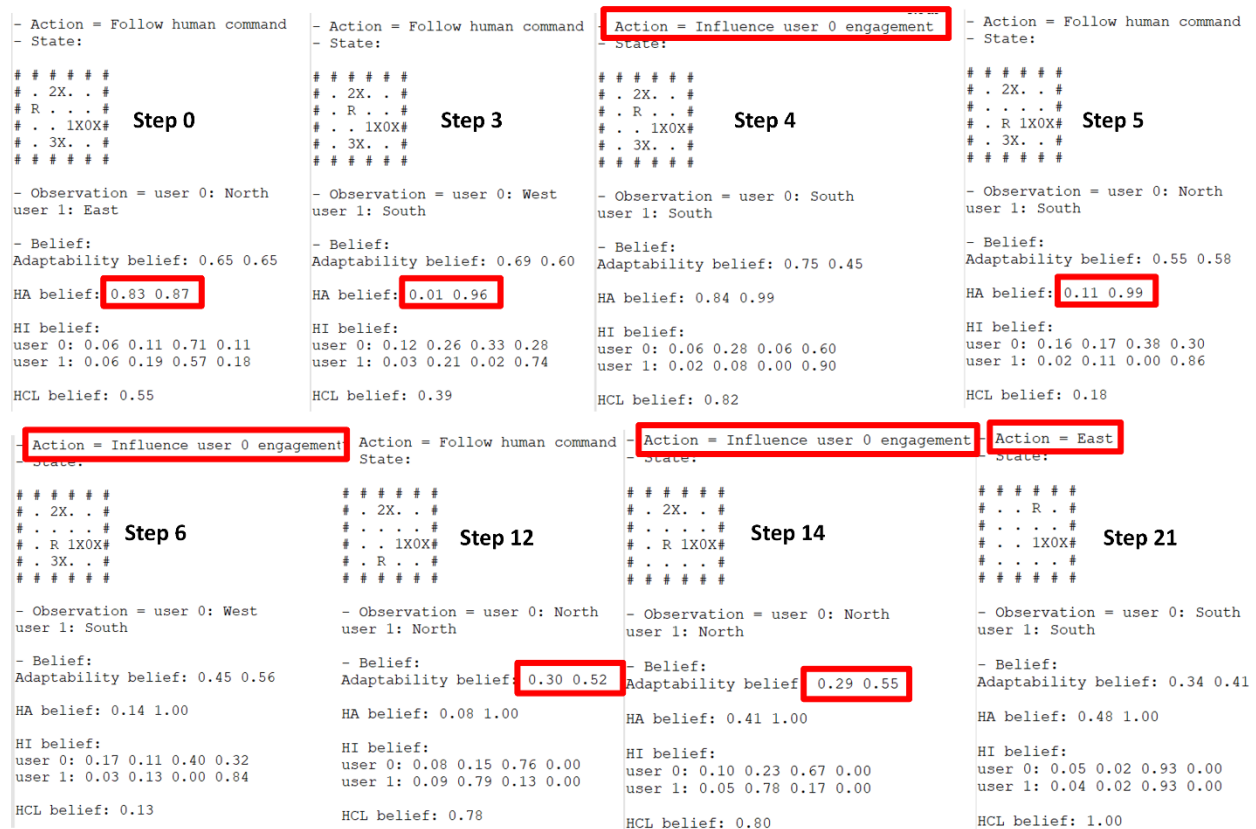


Figure 6-5. Case 3 Simulation Results.

## 6.6 Conclusion

The main contribution of this work is to design a novel formalized model of adaptation for multi-user HRI. To the best of my knowledge, although several researchers have proposed formalized models for one-to-one HRI, there is no mathematical models that formalize the interaction flow among human users and how robot actions could shape that interaction flow in the field of HRI. The success of such generalized models will pave the way for formal approaches to design and develop adaptive SAR systems for many-to-

one HRI that are applicable for people with special needs. I would also like to point out the limitations of this model. Compared to ROCARE, this model is less generalized but more rigorous. The task is more restricted due to the need of task formalization. However, this model provides a basis for a variation of formal models that adapt to each specific modification of the task formalization. For example, this particular task formalization requires HA and HCL for the entire HRI. In another set of tasks, maybe an adequate amount of HA and HCL is sufficient for task completion. This model is also limited in its sensing module. We only use users' interaction data to update HA, HI, and HCL. Potentially, the model could be expanded to incorporate other sensing elements such as engagement level (HA) measured by electrophysiological signals and HCL measured by mutual gaze. Despite these limitations, this is the first attempt to formally model multi-user HRI with integrated model of people, model of interaction, and model of machines. Simulation results demonstrated the ability of the model to estimate and shape HA, HI, and HCL, learn human adaptability, and facilitate task completion. In addition to simulation, user studies need to be conducted to test whether the simulation results are comparable to real HRI and assess the performance of the model.

## REFERENCES

- [1] A. Ahmad, G. Lawless, and P. Lima, "An online scalable approach to unified multirobot cooperative localization and object tracking," *IEEE transactions on robotics*, vol. 33, pp. 1184-1199, 2017.
- [2] T. Nestmeyer, P. R. Giordano, H. H. Bühlhoff, and A. Franchi, "Decentralized simultaneous multi-target exploration using a connected network of multiple robots," *Autonomous robots*, vol. 41, pp. 989-1011, 2017.
- [3] Y. Wang and I. Hussein, "Multiple vehicle Bayesian-based domain search with intermittent information sharing," in *Proceedings of the 2011 American Control Conference*, 2011, pp. 1280-1285.
- [4] Y. Wang, "Regret-based automated decision-making aids for domain search tasks using human-agent collaborative teams," *IEEE transactions on control systems technology*, vol. 24, pp. 1680-1695, 2016.
- [5] X. Wang, Z. Shi, F. Zhang, and Y. Wang, "Mutual trust based scheduling for (semi) autonomous multi-agent systems," in *2015 American Control Conference (ACC)*, 2015, pp. 459-464.
- [6] J. Y. Chen and M. J. Barnes, "Human-agent teaming for multirobot control: A review of human factors issues," *IEEE Transactions on Human-Machine Systems*, vol. 44, pp. 13-29, 2014.
- [7] E. de Visser and R. Parasuraman, "Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload," *Journal of Cognitive Engineering and Decision Making*, vol. 5, pp. 209-231, 2011.
- [8] G.-J. M. Kruijff, M. Janíček, S. Keshavdas, B. Larochele, H. Zender, N. J. Smets, T. Mioch, M. A. Neerincx, J. Diggelen, and F. Colas, "Experience in system design for human-robot teaming in urban search and rescue," in *Field and Service Robotics*, 2014, pp. 111-125.
- [9] P. Squire, G. Trafton, and R. Parasuraman, "Human control of multiple unmanned vehicles: effects of interface type on execution and task switching times," presented at the Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction, Salt Lake City, Utah, USA, 2006.
- [10] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, "Interactive robots as social partners and peer tutors for children: A field trial," *Human-Computer Interaction*, vol. 19, pp. 61-84, 2004.
- [11] Y. Matsusaka, H. Fujii, T. Okano, and I. Hara, "Health exercise demonstration robot TAIZO and effects of using voice command in robot-human collaborative demonstration," in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, 2009, pp. 472-477.
- [12] F. Faber, M. Bennewitz, C. Eppner, A. Gorog, C. Gonsior, D. Joho, M. Schreiber, and S. Behnke, "The humanoid museum tour guide Robotinho," in *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*, 2009, pp. 891-896.
- [13] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, "An affective guide robot in a shopping mall," in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, 2009, pp. 173-180.
- [14] W. Y. G. Louie, T. Vaquero, G. Nejat, and J. C. Beck, "An autonomous assistive robot for planning, scheduling and facilitating multi-user activities," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, 2014, pp. 5292-5298.
- [15] S. Keizer, M. E. Foster, A. Gaschler, M. Giuliani, A. Isard, and O. Lemon, "Handling uncertain input in multi-user human-robot interaction," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 2014, pp. 312-317.
- [16] Y. Kondo, K. Takemura, J. Takamatsu, and T. Ogasawara, "A gesture-centric android system for multi-party human-robot interaction," *Journal of Human-Robot Interaction*, vol. 2, pp. 133-151, 2013.
- [17] S. Chandra, P. Alves-Oliveira, S. Lemaignan, P. Sequeira, A. Paiva, and P. Dillenbourg, "Can a child feel responsible for another in the presence of a robot in a collaborative learning activity?,"



- in *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2015, pp. 167-172.
- [18] P. Alves-Oliveira, S. Janarthanam, A. Candeias, A. Deshmukh, T. Ribeiro, H. Hastie, A. Paiva, and R. Aylett, "Towards dialogue dimensions for a robotic tutor in collaborative learning scenarios," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 2014, pp. 862-867.
- [19] P. Alves-Oliveira, P. Sequeira, E. Di Tullio, S. Petisca, C. Guerra, F. S. Melo, and A. Paiva, "'It's Amazing, We Are All Feeling It!'—Emotional Climate as a Group-Level Emotional Expression in HRI," in *2015 AAAI Fall Symposium Series*, 2015.
- [20] I. Leite, M. McCoy, D. Ullman, N. Salomons, and B. Scassellati, "Comparing models of disengagement in individual and group interactions," in *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2015, pp. 99-105.
- [21] I. Leite, M. McCoy, M. Lohani, D. Ullman, N. Salomons, C. Stokes, S. Rivers, and B. Scassellati, "Emotional storytelling in the classroom: Individual versus group interaction between children and robots," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2015, pp. 75-82.
- [22] Z. Yumak, J. Ren, N. M. Thalmann, and J. Yuan, "Modelling multi-party interactions among virtual characters, robots, and humans," *Presence: Teleoperators and Virtual Environments*, vol. 23, pp. 172-190, 2014.
- [23] Y. Matsuyama, H. Taniyama, S. Fujie, and T. Kobayashi, "Framework of Communication Activation Robot Participating in Multiparty Conversation," in *AAAI Fall Symposium: Dialog with Robots*, 2010.
- [24] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro, "A two-month field trial in an elementary school for long-term human-robot interaction," *IEEE transactions on robotics*, vol. 23, pp. 962-971, 2007.
- [25] M. Vázquez, A. Steinfeld, and S. E. Hudson, "Maintaining awareness of the focus of attention of a conversation: A robot-centric reinforcement learning approach," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2016, pp. 36-43.
- [26] J. Fan, D. Bian, Z. Zheng, L. Beuscher, P. A. Newhouse, L. C. Mion, and N. Sarkar, "A Robotic Coach Architecture for Elder Care (ROCARE) based on multi-user engagement models," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, pp. 1153-1163, 2017.
- [27] S. Nikolaidis, Y. X. Zhu, D. Hsu, and S. Srinivasa, "Human-robot mutual adaptation in shared autonomy," in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2017, pp. 294-302.
- [28] S. Nikolaidis, S. Nath, A. D. Procaccia, and S. Srinivasa, "Game-theoretic modeling of human adaptation in human-robot collaboration," in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2017, pp. 323-331.
- [29] C. Baker, R. Saxe, and J. B. Tenenbaum, "Bayesian models of human action understanding," in *Advances in neural information processing systems*, 2006, pp. 99-106.
- [30] C. L. Baker and J. B. Tenenbaum, "Modeling human plan recognition using Bayesian theory of mind," *Plan, activity, and intent recognition: Theory and practice*, pp. 177-204, 2014.
- [31] C. L. Baker, N. D. Goodman, and J. B. Tenenbaum, "Theory-based social goal inference," in *Proceedings of the thirtieth annual conference of the cognitive science society*, 2008, pp. 1447-1452.
- [32] T. Ullman, C. Baker, O. Macindoe, O. Evans, N. Goodman, and J. B. Tenenbaum, "Help or hinder: Bayesian models of social goal inference," in *Advances in neural information processing systems*, 2009, pp. 1874-1882.
- [33] D. Sadigh, S. S. Sastry, S. A. Seshia, and A. Dragan, "Information gathering actions over human internal state," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 66-73.

- [34] S. C. Ong, S. W. Png, D. Hsu, and W. S. Lee, "Planning under uncertainty for robotic tasks with mixed observability," *The International Journal of Robotics Research*, vol. 29, pp. 1053-1068, 2010.
- [35] B. D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, J. A. Bagnell, M. Hebert, A. K. Dey, and S. Srinivasa, "Planning-based prediction for pedestrians," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, 2009, pp. 3931-3936.
- [36] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum Entropy Inverse Reinforcement Learning," in *AAAI*, 2008, pp. 1433-1438.
- [37] T. Smith and R. Simmons, "Heuristic search value iteration for POMDPs," in *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, 2004, pp. 520-527.
- [38] N. Ye, A. Somani, D. Hsu, and W. S. Lee, "Despot: Online pomdp planning with regularization," *Journal of Artificial Intelligence Research*, vol. 58, pp. 231-266, 2017.

## CHAPTER 7

### CONTRIBUTIONS AND FUTURE WORK

#### 7.1 Overall Contributions

The role of human-machine interaction (HMI) has been increasingly important in many aspects of our everyday lives. This dissertation focused on creating formal methods, algorithms, and architectures for adaptive HMI with specific applications to elder care and autism spectrum disorder (ASD) intervention. The main contributions of my research include: i) developing data-driven models of people to capture the mental states of the users; ii) designing and developing SAR systems that support both one-to-one interaction and many-to-one interaction for older adults in order to deliver multimodal therapies with an emphasis on social engagement; iii) designing model of interaction and model of machine for the control of SAR that take into account individual differences and are applicable for reuse across a variety of interaction scenarios; iv) designing and conducting pilot user studies to demonstrate the performance and acceptance of SAR systems by older adults in the laboratory setting; and v) conducting a multi-session field study and demonstrating, for the first time, the tolerance and acceptance of older adults residing at retirement communities and changes of their activity engagement and social interaction over time.

There is an urgent need for technological strategies to engage older adults in multimodal tasks that target physical, cognitive, and social functions and to provide treatments that are oriented to specific core symptoms and meaningful skills for individuals with ASD. The eventual success of a SAR system for people with special needs hinges upon three aspects: i) the system must be able to perceive and understand human behaviors and actions that are meaningful to the interaction or the task; ii) the system must be able to engage users in the interaction with the purpose of promoting meaningful changes in behavior or function; and iii) the system must know the relationship between users' responses and its own behaviors in order to provide appropriate feedback.

Most of the closed-loop SAR systems for people with special needs pay particular attention to the ability of the system to autonomously observe physical behaviors and actions of users. Human behaviors involve both explicit human behavior and the implicit mental states hidden behind behavioral performance. The mental state models of a user will allow the system to understand user's affective and cognitive states, task preference, and intention, which in turn could be used to adapt system behavior to make the system more aware of and responsive to users, and to make HMI more natural and efficient. In addition, the existing literature on SAR systems for older adults focuses particularly on one-to-one interaction to engage them in

a physical or cognitive activity. This limits the ability of SAR to provide multimodal therapies and to promote communication and social interaction by involving more than one older adult in the activity. Furthermore, interaction between a user and a SAR system is usually governed by a rule-based, task-specific model of interaction with robotic behaviors that are designed to be appropriate under certain user responses. This approach increases the dependence of SAR development on human experts by requiring complex handcrafted rules for the interaction design and the robot behavior design. This leads to time- and effort-intensive design of rules for each interaction scenario. As the complexity of the interaction increases, it becomes more difficult to design handcrafted rules that could accommodate for individual differences and thus would increase the likelihood of administrator intervention during HMI. Finally, the feasibility of SAR systems needs to be evaluated with target population not only in the laboratory setting, but also in the field. Only a few researchers have evaluated the performance and user acceptance of more advanced SAR systems in the field. This dissertation research aimed to address these problems.

## **7.2 Technical Contributions**

### **7.2.1 Models of people, interaction, and machine**

The first set of technical contributions lies in the design and development of various models for intelligent HMI. Data-driven mental state models of individuals with ASD were built based on their implicit or passive electroencephalogram (EEG) data during HMI. Previous literature has explored physiological signal, eye gaze data, body posture, tactile event, and verbal intonation for mental states estimation. Despite the increasing interest in developing EEG-based passive brain-computer interface applications to enrich HMI, there is a paucity of research on EEG-based affective states recognition for real world tasks such as driving and lack of models for individuals with ASD, whose EEG activity is different from their typically developed peers. We built group-level classification models that were capable of recognizing binary low and high intensity of four affective states, including engagement, enjoyment, boredom, and frustration, as well as mental workload of individuals with ASD in the context of driving. Results implied that models based on EEG activations can detect with high accuracy the states of low engagement, low enjoyment, high frustration, and high workload for ASD population. Boredom recognition had relatively low accuracy. The primary technical contributions of this work are: i) integration of an EEG sensory modality into a virtual reality-based driving system to collect EEG data from individuals with ASD during realistic driving tasks; ii) development of a two-step feature calibration method to allow for group-level training. This dramatically reduces the training sessions needed compared to individualized model training; iii) systematic evaluation of feature generation approaches to demonstrate the possibility of group-level affect and workload

recognition based on EEG data; and iv) systematic evaluation of feature and electrode usage to identify discriminative features.

In CHAPTER 3, we have designed a SAR architecture for one-to-one and many-to-one HRI. Most systems to date have predominantly focused on one-to-one human-robot interaction (HRI) and the models of interaction are task-specific. We designed a novel multi-user engagement-based robotic coach system architecture (ROCARE) with four main features: i) mathematical models for each module and relationships among modules instead of simple interconnection; ii) engagement models to capture the dynamics of human-human interaction (HHI) and HRI; iii) capacity for both one-to-one interaction and many-to-one interaction; and iv) generalizability of the architecture for different HRI scenarios. Based on ROCARE, we implemented two SAR systems and conducted two preliminary feasibility studies, a one-to-one interaction with five activities and a triadic interaction with a single activity. EEG signals were analyzed offline to evaluate the engagement level of older adults in the robot-mediated activities. The results indicated that i) ROCARE was positively accepted by older adults with and without cognitive impairment; ii) ROCARE can be used for one-to-one and multi-user HRI; and iii) our selection of the EEG feature has the potential for objectively measuring older adults' engagement intention and harnessing it to realize individualized activity management.

ROCARE was designed to focus on user engagement and the effect of user engagement on activity management. This property makes ROCARE a model for long-term engagement. In terms of robot behavior adaptation within one activity, ROCARE defined system goal for a robot as a function of users' HRI engagement and/or HHI engagement. Within one activity, robot behaviors were defined to be controlled by a reactive model or through learning algorithms and conditioned on the task difficulty level, the engagement variables, the interaction history, and the system goal. Although we have defined the relationship between the robot behavior and the interaction together with the user engagement model, the actual model of adaptation that governs the robot behaviors within one activity was left to be designed based on the needs of a particular HRI scenario. In order to further reduce the time- and effort-intensive design of model of adaptation for each interaction scenario, we formally modelled multi-user HRI with well-defined task formulation and a mathematical model of people, interaction, robot, and their integration. The main contributions of this work are: i) the problem formulation of a cluster of multi-user HRI scenarios that is specifically designed for SAR applications with an emphasis on enforcing HHI through HRI; and ii) the mathematical model for multi-user HRI that formally defined and integrated a model of user, a model of robot, and a model of interaction.

### 7.2.2 SAR systems development

The second set of technical contributions is in the design and development of SAR systems for older adults to deliver multimodal therapies and to foster social interaction. Three SAR systems were developed based on ROCARE to administer one-to-one and triadic interaction with older adults. The initial work focused on engaging older adults in multi-modal interaction during which we developed physical activities and cognitive activities. Specifically, a semi-autonomous SAR system was developed to administer five activities that were both passive and active, which were an orientation task (active), solving simple math problems (active), observing the robot dance to music (passive), a form of the “21 questions” game where the robot determines the person’s birth state (active), and a joint chair exercise (active). Prompts and reinforcements were developed and embedded in the system. A Kinect RGBD sensor detected in real time the gestures of the participants during the chair exercise activity. EEG and galvanic skin response were continuously recorded to evaluate the engagement level of the participants in the robot-mediated activities. User study results indicated that our SAR system was well tolerated by older adults, and they were interested and engaged in these activities.

The second SAR system extended the chair exercise activity from one-to-one interaction to triadic interaction. A fully autonomous SAR system was developed to administer a gesture-based imitation game. The finite state machine-based gesture recognition algorithm was scaled up to detect simultaneously two exercise motions performed by older adults. Prompts and feedback were embedded to cue each individual or the pair as a whole. Participants’ eye gaze data were estimated by their head pose angles detected by a Kinect sensor and were logged during HRI to evaluate their engagement towards each other. Older adults had positive perceptions on triadic HRI after the experiment. Social communication between pairs of participants could be elicited by the robot as seen from both video recordings and head pose data. This work indicated the potential for SAR systems to involve more than one person in the hope that such many-to-one interaction would facilitate some interpersonal communication.

The acceptability and initial results from the previous two SAR systems were encouraging, which motivated further development of SAR-Connect, an autonomous robot-mediated interaction system to foster social interaction among older adults within a multimodal task. SAR systems in existing literature focus on tasks that have a single modality, cognitive or physical. This system consisted of three major components, which were i) a multimodal task with embedded physical, cognitive, and social stimuli; ii) a robot control mechanism to keep older adults engaged in both HRI and HHI; and iii) data analysis algorithms to quantify older adults’ social interaction and activity engagement. We designed a motion-based user interface by means of a Kinect sensor to involve older adults’ in physical movement, and developed a virtual book sorting task to involve older adults’ in cognitive activity. For social engagement,

we designed collaborative rules to encourage social communication and collaboration between older adults. The system was modeled by timed automata and hierarchical state machines (HSM) to support both one-to-one interaction and triadic interaction, to keep older adults physically and cognitively engaged, and more importantly, to foster interpersonal social interaction between older adults themselves. To capture older adults' social interaction and activity engagement, we developed a suite of data analysis algorithms to quantify HRI and HHI from multiple sensory modalities, including game interaction data, head pose, vocal sound, and EEG. Results from a user study showed that SAR-Connect could involve two older adults to perform multimodal activities, could engage them in HRI and HHI, and could quantitatively measure their social interaction and activity engagement.

### **7.2.3 Automatic evaluation methods**

SAR systems need to achieve measurable progress, which includes progress on physical and cognitive functions, activity engagement, and social interaction. Unlike traditional robotic systems or personal service robots, the progress of older adults cannot be simply extracted from the task specification. Although robot behaviors are tailored to older adults' task performance, performance itself is not a good indicator of older adults' progress due to their vulnerability and the nature of aging. On the other hand, manual analysis of older adults' behaviors by a trained human rater is effort and resource intensive. These require methods to automatically evaluate the progress of older adults during robot-mediated interaction. In CHAPTER 3, we computed the EEG engagement index (EEI) to estimate older adults' engagement level during HRI. EEI was validated with participants' own rating of activity preference on a 5-point Likert scale (correlation: 0.73,  $p < 0.001$ ). In CHAPTER 4, we developed algorithms to automatically evaluate older adults' activity engagement and social interaction during HRI from three types of data, which were game interaction, head pose, and vocal sound. The head pose and vocal sound analysis algorithms were validated by comparing algorithm detection results to manual video analysis results by a trained human rater. The head pose analysis algorithm could detect with high accuracy the amount of times (precision: 98.3%, recall: 91.5%) and the number of times (precision: 95.7%, recall: 86.3%) older adults looked towards their peers. The vocal sound analysis algorithm could detect with high accuracy the number of times (precision: 97.9%, recall: 87.0%) older adults spoke. The algorithms developed in CHAPTER 4 were then used to assess changes of older adults' activity and social engagement in the field study.

### 7.3 Contributions to the Science of Elder Care and ASD Intervention

Besides technical contributions, the research presented in this dissertation also contributes towards the science of elder care by providing controllable SAR systems to engage older adults in multimodal activity-oriented therapies and to engage more than one individual simultaneously with mechanisms to foster HHI in order to alleviate social isolation and/or loneliness in older adults. In addition, the developed data analysis algorithms provide a more efficient way to quantify activity engagement and social interaction, which will reduce the burden on resources to manually analyze and code human behavior. Such technologically sophisticated systems are expected to play an important role in addressing the lack of healthcare resources, enhancing older adult's function and quality of life, and reducing burden on the caregivers. The dissertation research provides important insights into the development of SAR systems to provide meaningful activities for older adults and the feasibility of applying such systems in the real world.

As part of the dissertation research, we designed and conducted a set of user studies with target population to test our SAR systems. In CHAPTER 3, we first conducted an initial alpha testing of ROCARE with 11 older adults of which 4 had preexisting diagnosis of mild cognitive impairment or dementia. We then recruited 14 older adults who were paired for simultaneous interaction with the robot. In addition to system development and user study design, we developed a robot acceptance scale (RAS) to measure older adults' perceptions about interacting with a robot. We found that i) ROCARE could be used with older adults; ii) the robotic platform and interaction tasks, one-to-one interaction as well as triadic interaction, could be engaging to older adults; iii) RAS could measure older adults' perceptions about interacting with a robot; and iv) older adults would interact with each other with the aid of the robot.

In CHAPTER 4, the SAR-Connect system was tested by 26 older adults, of which 18 completed both one-to-one and triadic sessions. This study indicated the potential to use SAR-Connect to involve two older adults to perform activities together with the aid of an intelligent system and quantitatively measure their activity. Specifically, we found that i) older adults would engage in a virtual reality-based physically, cognitively, and socially stimulating activity; ii) SAR-Connect could be potentially useful to foster social interaction in addition to activity engagement; and iii) objective measurement from the system could be used to evaluate older adults' social and activity engagement. We believe that such a system will be helpful in providing both physical and cognitive activities to older adults to keep them engaged, foster interpersonal interaction beyond the interaction with the robot and quantify their interaction.

In order to test the feasibility of robot-mediated triadic interaction in the real world setting, the two SAR systems for triadic interaction developed in CHAPTER 3 and CHAPTER 4 were integrated to create Ro-Tri (CHAPTER 5). We performed a multi-session field study with older adults residing at two local



retirement communities and demonstrated, for the first time, the tolerance and acceptance of older adults with or without cognitive impairment and changes of their activity engagement and social interaction over time. Findings from the study indicated slightly improved social interaction during HRI in terms of visual attention towards other older adults and maintained or slightly improved interests, perceptions on the system, and engagement in HRI.

Finally, the work described in CHAPTER 2 contributes towards the science of ASD intervention by building affective state and mental workload models of individuals with ASD for potential affect- and workload-based individualized driving skill training intervention. We believe that the ability to tailor individual learning experiences based on their affective states and mental workload will make the skill training more effective. This work provides a proof of concept that such EEG-based recognition could be useful to individualize ASD intervention. Although the data-driven models were developed specifically for the driving tasks, the approach as well as feature and electrode selection results are transferable to personalize other advanced training systems developed for ASD intervention.

## **7.4 Future Work**

One of the future directions of this research is to further improve the models and continue integrating them into SAR systems and evaluating their performance with the target populations. Evidence-based intervention needs to be developed for personalized driving skill training of individuals with ASD by incorporating the predictions from the affective state and mental workload models to the intervention paradigm of the driving system. Another area of emphasis in future could be integrating various mental state models of people, such as affective state, mental workload, intention, and attention estimation, to allow SAR systems to adapt under more sophisticated and subtle situations and make the HRI more natural and efficient. As more activities are developed for older adults, the performance of activity management module in ROCARE could be assessed through a multi-session user study. The mathematical model we have designed is the first attempt to formally model multi-user HRI with integrated model of people, model of interaction, and model of machines. The applicability of such a model need to be tested by several realistic tasks for people with special needs and assessed by responses of the target population. Furthermore, it would be useful to expand the model to incorporate other sensing elements such as engagement level measured by EEG and HHI measured by mutual gaze.

In this dissertation research, we have developed three SAR systems to interact with older adults through speech and gestures, provide meaningful feedback/instructions, and encourage communication between two older adults. We have also developed algorithms to detect task action, speech duration, gaze

direction, and gesture and to estimate older adults' activity engagement and social interaction. These systems could be further strengthened in several ways. First, the quantitative data from each individual sensing module could be fused to generate a more robust estimation of older adults' activity engagement and social interaction. Second, the current systems provide feedback based only on older adults' task performance. By adapting robot behaviors based on the estimation of engagement and social interaction in addition to task performance, it would allow richer and more meaningful HRI. Third, an intention detection module could be added to keep track of the interaction data and predict user intention for the purpose of providing user-specific feedback and user-specific task profile.

It would also be of great importance to make the SAR systems more robust such that non-experts can operate it and to expand the library of tasks to address varying degrees of cognitive and physical impairments of older adults. The field study conducted in this dissertation research demonstrated initial feasibility results of Ro-Tri to engage older adults in HRI as well as HHI over time in a real world setting. In order to systematically examine responsiveness and engagement among older adults with cognitive impairment as well as effect on cognitive, physical, and social function, SAR-based intervention for older adults needs to be explored in a clinical trial that systematically i) matches participant's capability with the sophistication of the SAR systems; and ii) creates tasks that combine physical, cognitive, and social elements.