Multimodal Collaborative Virtual Environment with Embedded Intelligent Agent to Facilitate Teamwork and Executive Functioning: Applications for Individuals with Autism

By

Ashwaq Zaini Amat Haji Anwar

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
Vanderbilt University
May 12, 2023
Nashville, Tennessee

Approved:
Nilanjan Sarkar, Ph.D.
Gabor Karsai, Ph.D.
Keivan Stassun, Ph.D.
Amy S. Weitlauf. Ph.D.
D. Mitchell Wilkes, Ph.D.

ACKNOWLEDGEMENTS

I would like to sincerely thank my advisor, Professor Nilanjan Sarkar for his wisdom and invaluable patience throughout my PhD journey. He is always there to give the right advice at the right time, to give you the push that you need, and most importantly to guide you to be a good researcher. I would like to also extend my gratitude to Dr. Amy Weitlauf who has been providing valuable feedback and knowledge on autism interventions that helped tremendously in my research completion. To Dr. Mitch Wilkes, thank you for entertaining my questions on HMM and K-means, to drive the solution of the probabilistic model. I want to also thank other members of my PhD committee, Dr. Keivan Stassun, Dr. Gabor Karsai, for their insights and encouragements to finishing this dissertation.

To my colleagues and fellow researchers at Robotics and Autonomous Systems Lab, thank you for being who you are. I am very grateful to have had the opportunity to work directly with Deeksha, Mahrukh, and Abigale across various parts of the research. They all brought their A game and contributed significantly to the dissertation. Thank you Mira for all the support and being the best lab manager. Thank you Akshith, Alexandra, Ritam, and Nibraas. My experiments would not have been successful without the help from Ms. Amy Swanson. She put in a lot of effort to schedule participants for the experiments. To Kelley, thank you for sorting out all the compensations for the experiments. A special thanks to Professor Medha Sarkar who provided me the opportunity to mentor her students developing useful projects. Both Medha and Nilanjan showed me that no matter what you are facing in life, keep on going. You both have so much to give, and I will treasure and put it to practice in my life too.

I am forever indebted to my husband, Hubab Asmawi and our three children, Rayyan, Muslih and Hannah, for holding my hand throughout this journey for the past six years in a foreign land, so that I could earn my PhD. To my family members and friends, thank you for being a part of my life that I will continue to appreciate every day. The highest praise and thank you to God Almighty for the countless blessings to finishing this journey.

Table of Contents

List of Figures

List of Tables

CHAPTER 1: INTRODUCTION

This chapter will discuss the background of my work and describe my research goals. In the background, I will first briefly describe the importance of teamwork and executive functioning skills, followed by how HCI-based systems can be used as training tools for these skills, and finally briefly describe the application of HCI-based systems that could facilitate individuals with disabilities to learn these skills. I will be using both identity-first and people-first language to respect both views by interchangeably using the term ‘autistic individuals’ and ‘individuals with ASD’ [1].

1.1 Background

Teamwork and executive functioning (EF) are examples of skills needed for success in the future [2]. According to a report led by Microsoft Corporation, the capability to communicate and collaborate with colleagues (teamwork) and capability in time management, planning, and critical thinking (executive functions) are among the core skills that they are looking at in future employees [3]. Human-computer interaction (HCI) technology such as digital games has been shown to positively impact the training of these skills [4]. Minecraft is an example of a popular digital game widely used in classroom settings to teach a wide range of skills due to its flexibility, ease of access, and engaging appeal [4]. However, commercial digital games lack the structure to scaffold skill learning, do not provide real-time feedback or prompts that could facilitate skill learning, and have no objective means of measuring players’ skills improvements. A collaborative virtual environment (CVE)-based system can be created to address the limitations of conventional digital games as mentioned above. A CVE-based system can be designed with explicit learning objectives as a serious game, provide real-time individualized feedback and prompts, and capture quantitative measures that can facilitate skill learning and assessment. A growing body of research explores the application of serious games for individuals with disabilities to complement existing interventions and learning. Individuals with disabilities, particularly Autism Spectrum Disorder (ASD), have been reported to have diminished social communication and cognitive skills, which are core components of teamwork and EF. However, to design and develop a HCI-based skill learning tool, we need to understand how technology can be used to support the learning process [5]. My research goals will attempt to address these research questions through the development of a novel CVE with team-based activities to support teamwork and EF skills training, with an initial application for autistic individuals.

The rest of this chapter is organized as follows. Section 1.2 discusses my research goals based on existing literature related to HCI-based systems with individuals with ASD, primarily CVE systems that attempt to support social skills development in individuals with ASD, existing methods of evaluating and measuring teamwork, and EF in individuals with ASD, and intelligent agent design within CVE-based systems. Section 1.3 reviews different inclusive design considerations for digital games for ASD interventions and their importance. Section 1.4 addresses the opportunities for research contributions. Section 1.5 summarizes my research work.

1.2 Research Goals

Teamwork can be defined as the act of working together in a collaborative manner to achieve a common goal efficiently [6]. Executive functioning is a set of skills that requires cognitive control and behavioral competencies such as self-regulation, working memory, and cognitive flexibility [4]. Both skills are important throughout a person's life and often go together [3]. Interest in teamwork-related research has increased rapidly over the last decade, focusing primarily on the influence of teamwork in organizational and workplace success [8]. Effective teamwork skills training can improve team performance and quality of the skills [6]. According to Salas et al., simulation-based training (SBT) is an effective teamwork training tool as it enables individuals in the team to engage in a shared social, cognitive, and behavioral process pertaining to teamwork while receiving feedback based on their performance [7]. However, studies that investigate teamwork and EF skills together are limited and effects of training these skills together remain underexplored. Serious games have been shown to positively benefit training of teamwork and EF skills [4]. Simulated serious games allow users to develop various teamwork and EF skills by experiencing and applying the skills in an environment that represents a real-world scenario [9]. Some common use of simulated serious games in skills training includes firefighter training [10], surgical team teamwork training [11], and vocational training [12].

Autism Spectrum Disorders (ASD) is a developmental condition that can be characterized by challenges in social interaction, communication and restrictive or repetitive behaviors [13]. According to the Centers for Disease Control and Prevention (CDC), 1 in 44 children are diagnosed with ASD each year based on 2018 data [14]. In another report, approximately 70,700 children with ASD reached adulthood each year [15]. CDC also estimated that 2.21% or 1 in 45 adults in the United States are diagnosed with ASD [16]. One major challenge for autistic adults is securing and retaining employment. Employment is a major contributor to the quality of life of adults and remains one of the main challenges faced by individuals with ASD as they transition to adulthood. Compared to other individuals with disabilities, adults with ASD have the highest unemployment rate between 50 – 85% [17]. For those with employment, the majority of them are either underemployed or unable to retain their position [18]. It is a fact that many individuals with ASD are highly skillful [19], have great attention to detail [20], and do not mind repetitive jobs [21], which are qualities sought after by employers. However, core deficits in social communication and interaction skills have cast a shadow against the outstanding qualities they possess [22]. Deficits and delays in developing these skills may hinder the development of more complex social skills such as teamwork and EF. Studies have shown that teamwork gives individuals with ASD the opportunity to build upon their social communication skills [23], problem-solving skills [24], and self-confidence [25]. Although existing training and interventions have shown some improvements in teamwork skills in adolescents with ASD, simulating real-world teamwork scenarios can be tedious, resource-straining, and costly, thus limiting the accessibility and reach of the interventions. Given these circumstances, current intervention and vocational practices to prepare adults with ASD for employment need to be improved.

Over the last decade, the use of HCI technology has shown promising benefits that can potentially complement conventional ASD interventions by providing engaging interactions and replicable solutions that can minimize costs and provide relatively broader access to users [26]. Virtual reality (VR) can be used to simulate teamwork activities based on real-world scenarios at a lower cost. Other characteristics of VR such as consistency, interactivity, and predictability are appealing for autistic individuals as they have a natural affinity for technology-based interactions and prefer stability [27]. Some examples of VR-based systems for ASD interventions are focused on teaching both social skills and technical skills, which include skills such as cooking [28], road safety [29], driving [30], joint attention [31], and emotion recognition [32]. Generally, these studies reported that autistic individuals were able to interact positively in the virtual environment and showed potential improvements in their skills and positive behaviors after using the systems. Nonetheless, conventional VR-based systems are limited to single user interaction and unable to support complex back and forth human-human interactions, which is important for training teamwork skills. Additionally, when considering conventional VR-based interactions for individuals with ASD, they might be more comfortable interacting with a virtual avatar compared to a human partner, and thus making it less efficient for generalization to real-world [33 - 35]. A collaborative virtual environment (CVE), alternatively, can support multi-user interactions, allowing users to communicate with each other in a natural manner while performing a task together in the shared virtual environment. At the same time, CVE can potentially reduce the attachment effect of a conventional VR-based interaction in individuals with ASD. As such, the first goal of this dissertation is to design and develop team-based activities in a CVE as a skill training tool that can support the development of teamwork and EF skills. While the work were demonstrated with autistic individuals, the design principles and application are suitable for other populations as well.

Vocational and technical skills are important aspects of employment and are the main criteria considered for employment [36]. However, interpersonal or professional skills such as teamwork and EF are the core skills needed to be able to secure and retain employment [37]. Studies have shown that skills such as teamwork and EF can contribute to improved productivity and workplace performance in a shorter time [38, 39]. Teamwork is also one of the seven professional skills recommended by the Department of Labor for youth with disabilities preparing for employment [40]. The importance of teamwork and EF are also reflected in the hiring process of companies like Microsoft and Specialsterne, which employ autistic individuals. They use a non-traditional interview process for autistic candidates to assess teamwork and executive function skills and use tasks such as Lego Mindstorm group projects [41] and Minecraft collaborative tasks [42] for evaluation. However, evaluation of such skills remained subjective and prone to bias. Among the studies that do investigate social skills improvements, they mostly rely on qualitative measures of performance and questionnaire methodology. Teamwork is a complex social skill as it requires verbal, non-verbal, and physical cooperation and coordination between two or more people. For instance, when two people are trying to move a heavy object from one location to another, they would need to: 1) effectively communicate how and when they will move it (verbal and non-verbal), then 2) coordinate their movement when picking up the object (non-verbal, physical), and 3) move in the same direction (physical), and d) use about the same amount of force to carry the object (physical). Similarly, executive functioning skills are also complex to measure. For instance, to measure time management skill, it is not sufficient to only measure it based on the ability of the users to finish the task on time. It is also valuable to identify if users were aware of the time and how they effectively plan and use the time by working together. Thus, there are potential benefits of capturing interaction data from multiple modalities to evaluate these complex skills. Multimodal analytics is an emerging area of research within HCI that concerns the analysis of integrated data from various sources to identify measurable parameters that can be used to evaluate the skills and provide researchers with an extensive understanding of the skills being learned [43]. Applying multimodal analytics to quantify various dimensions of collaboration skill is one way to provide a consistent evaluation of the skill. Existing studies related to collaborative learning have identified and defined several dimensions that are relevant for evaluating teamwork and EF [44 - 47]. Incorporating multimodal measures into the defined dimensions allowed us to objectively assess these skills, enabling us to understand how the skills can be learned and ways we can support the skills development. Additionally, these measures served as a more accurate and reliable evaluation of skills. Therefore, another goal of this research is to capture multimodal inputs from various input devices and integrated sensors and identify which data can be used to represent different dimensions of teamwork and EF skills and assess the performance of these skills for both users in a CVE system.

Although CVE-based systems have the potential to support training and development of social interaction and communication skills in a virtual setting, they are still limited by the lack of flexible and individualized responses [48]. An intelligent agent can be implemented in CVE-based systems to facilitate and improve learning experience in a collaborative virtual setting and has been widely explored in the area of HCI research [49]. An intelligent agent has the capability to perceive user’s behavior and generate individualized responses based on the need of the user [50]. Additionally, intelligent agents have primarily been used in single-user systems, where the agent takes the role of an interaction partner as part of human-agent interaction to converse and perform actions with the user either by turn-taking or joint action.

However, an intelligent agent in a single-user system has restricted interaction capability since the agent is usually implemented using a simple rule-based model or pre-defined responses that can only perceive limited types of interaction behavior and will not be sufficient to support the complex nature of human-human interaction. Moreover, human-agent interaction in social communication studies removes the element of real-world human-human interaction and might hinder developing social skills required for interaction with another person. Research has pointed out challenges of transferring the learned skill in a human-agent interaction to real-world human-human interaction [51]. In recent years, researchers have been exploring the use of probabilistic modeling and machine learning (ML) in the design of the intelligent agent in collaborative interactions as a possible solution to process the complex behavior of human interaction. Instead of using a fixed value or rule to perceive and respond to human action, a probabilistic model adds flexibility by analyzing and training user data to represent a wider range of possible perceived actions. Studies have reported positive outcomes of using probabilistic model and ML in intelligent agent design [52, 53, 54]. An intelligent agent that incorporates a flexible prediction model has the potential to provide responses that are more specific and suitable to the user, thus improving the user's learning experience. Currently, intelligent agent studies in multi-user collaborative systems mainly focus on moderating and delivering collaborative learning content rather than supporting skills development [55]. Studies that implement an intelligent agent to facilitate teamwork skills learning remain underexplored. Thus, another goal of this research is to design and embed an intelligent agent in a CVE system that can monitor and facilitate human-human interaction through a reliable prediction model and efficient feedback mechanism.

The research presented here focuses on the design, development, and application of team-based virtual collaborative activities and an embedded intelligent agent within a CVE system designed to predict users' collaboration state and assess collaboration performance. The research aims include: 1) the design and development of various team-based virtual tasks in a CVE that can encourage and foster teamwork and executive function skills in peer-based interactions, 2) the development of a multimodal data analytics model that can provide an understanding of teamwork and EF performance from quantitative data, and 3) build an intelligent agent that can predict user's collaboration states using a probabilistic model to provide flexible and individualized feedback to the users that can facilitate the development of teamwork skills. This research produces a novel virtual team-based activity simulator that can enable individuals to practice their teamwork and executive function skills. The application of the system is focused on individuals with ASD with the hope of improving their employment landscape and expanding the accessibility of ASD interventions.

1.3 HCI-based Serious Games to Support Skills Development and Training

The use of digital games as a tool to teach specific skills has increased rapidly over the past decade [4]. One form of digital game, called serious game, is created for training and imparting skills with specific learning objectives, which differentiate them from regular digital games. In recent years, serious games have been designed for individuals with disabilities, including autistic individuals [56]. Serious games are designed with explicit learning objectives embedded within interactive and entertaining games that can be appealing to individuals with ASD. Primarily, serious games provide an interactive learning environment with collaborative activities that keep users engaged and motivated [57]. VR technology offers potential benefits as a serious game platform. VR-based systems have the flexibility to simulate a real-world scenario that can either be complex or simplified based on the need for training and allowing users to learn by experiencing the lesson themselves. For example, studies have explored VR-based systems to teach skills ranging from customer service skills [81] to driving skills [30]. However, collaborative interactions in VR cannot support the natural and complex interactions that characterize real-world interactions. Instead, CVE possesses the same characteristics of VR with the added advantage of allowing multiple users to interact with each other while playing serious games, creating an unrestricted collaborative interaction similar to real-world interactions. CVE is a multimodal interactive technology that can be integrated with various devices and sensors such as eye trackers, haptic devices, and microphones, enabling users to easily experience and interact with the simulated environment. These devices and sensors provide researchers with rich quantitative data to better understand teamwork behaviors in a simulated environment.

Advancement in machine learning (ML) and artificial intelligence (AI) has enabled complex data analysis that can be used to predict and assess users' actions more accurately. With the availability of such information, an intelligent agent can be designed to provide individualized responses that can encourage and scaffold new skills development while keeping users engaged throughout the interaction. Individualized interventions can target the needs of individuals with ASD as they have unique challenges and strengths. Overall, VR-based technology is a suitable platform for developing serious games that aim to train specific skills, including social behavior skills for individuals with ASD.

This work aims to design virtual team-based activities in CVE embedded with an intelligent agent to facilitate teamwork and EF skills training and assessment, with potential benefits for individuals with ASD. Hence, the literature reviews discussed in the following subsections are divided into three main focus areas on technologies and research related to our work. First, we discuss the existing VR-based and CVE-based serious games used for learning new skills or improving existing skills. Next, we present studies employing intelligent agents in HCI-based interaction systems and different methods to perceive human behavior. Finally, we discuss studies that involved multimodal data acquisition and analytics from single and collaborative interactions.

1.3.1 HCI technology to support skill learning

Research in the field of HCI mainly investigates ways computer interactions can benefit and enhance human experience and improve quality of life [5, 58]. In this area, many studies investigated the use of HCI systems in improving human learning experience [59, 60]. For example, in 1973, Colby developed a computer system that displayed a symbol on screen accompanied by a voice and other sounds for non-verbal autistic children [61]. Results showed linguistic improvements in 13 out of 17 children with ASD. The remaining 4 children did not complete the task and did not show any improvements in their skills. Similar to this study, other studies have also explored the use of HCI-based systems to train and teach a wide range of other skills that encompass cognitive skills; visual-spatial, auditory, and speaking skills [62], recall, attention, and concentration [63], affective learning such as emotion regulation [64], motor skills [65, 66, 67], and communication skills [68, 69]. Conventional HCI systems can be as simple as a display unit with a keyboard and mouse as input devices. More advanced systems may incorporate additional interaction modalities such as a microphone for speech input, an eye tracker for gaze input, and a gamepad controller that can provide more range of motion than keyboard and mouse. The additional interaction modalities immerse users in computer interactions that can closely replicate real-life interactions, possibly transferring these skills to real-life [70 -72].

1.3.2 VR-based systems as a skill learning tool

VR is an example of a HCI technology that can support multimodal interactions and it has additional features that make it suitable for learning. VR-based systems can simulate a safe and controllable environment for learning. For example, VR-based systems have been used to train firefighters [73, 74], and provide construction safety training [75]. By controlling the complexity of the virtual environment, many VR-based systems have also been used to train individuals with disabilities, including ASD. For example, Saiano et al. designed a VR-based system to teach adults with ASD road safety skills. Participants attended ten training sessions with increasing complexity in each visit, giving participants room to develop their skills [29]. By the end of the training sessions, participants demonstrated improved performance in street-crossing skills. Zheng et al. created a personal hygiene tutorial for children with ASD, teaching tooth brushing skills step-by-step [76]. The study reported that participants were less stressed and had improved tooth brushing performance. VR interactions can provide real-time prompts and feedback through visual, audio, and tactile prompts, improving user experience and engagement. In the same study mentioned above, the researchers designed a feedback mechanism that provided hints to the children when they were not doing the task correctly and positive encouragement when they did the task correctly [76]. Another feature of VR that makes it suitable for learning and training is the ease of development and deployment of the technology, compared to traditional classroom training sessions. This can be seen in the shift of employee training methods used by big corporations such as Walmart and BMW, where they are slowly integrating the use of VR technology to train their new employees [3, 77, 78]. VR-based systems hold great potential as a learning tool for specific skills based on wide range of skills and learning benefits that can be achieved through these systems.

1.3.3 CVE-based systems as a skill learning tool

Social interaction and communication are essential life skills that can bring life-long benefits. Conventional VR-based systems have been used to help individuals train and improve their social interaction and communication skills, such as public speaking skills [79] and interviewing skills [80]. Bozgeyikli et al. developed an immersive VR-based system for individuals with disabilities to assess and train vocational and social skills [81]. Results showed that participants could interactively practice their technical skills within the virtual environment. Still, social interactions were limited to basic question-answer scenarios not reflective of real-world interactions. As can be seen, conventional VR-based systems have restricted communication capabilities based on limited or pre-recorded responses. They are unable to support the more complex social interactions such as negotiation, collaborative discussion, and teamwork. A CVE extends conventional VR technology by retaining the same capabilities that VR-based systems can offer, with the added advantage of supporting multiple users within the same shared virtual space, allowing users to interact and converse with each other freely. CVE-based systems have been primarily studied to understand the impact of collaborative learning. For example, Breland and Shiratuddin designed a CVE-based system to investigate how well architecture students work together in a shared virtual space compared to working individually in a virtual environment [82]. The authors reported that the students had improved productivity, made less design errors, reduced stress, and increased positive mindset when working in a collaborative setting. Another study designed a collaborative game for school-aged children to enhance learning and collaborative skills [83]. The result from a pilot study indicated children enjoyed collaborating with partners when they were learning new skills.

CVE-based systems are also actively explored as social skills intervention tools for individuals with ASD. iSocial, a distributed learning environment, was designed for school-aged children with ASD on an existing 3D virtual platform, Second Life [84]. Eleven children with ASD were trained in three areas of social cognitive processes, which were Theory of Mind (ToM), emotion recognition, and executive functioning. Results demonstrated promising learning outcomes in social competencies for the children. In another study using Second Life, Kandalaft et al. designed a series of social scenarios as training tasks for individuals with ASD [85]. Eight adults with ASD were recruited and completed 10 training sessions. As part of the training, an individual with ASD interacted with two clinicians in the virtual environment. The clinicians provided the autistic individuals with instructions and prompts while they navigated through the scenarios in the CVE. The study reported three improvement measures: verbal and non-verbal emotion recognition, ToM, and conversational skills. Participants showed significant improvement in verbal and non-verbal emotion recognition and ToM, while conversational skills showed improvement, but were not significant. In a more recent study, Vona and colleagues designed Social MatchUP, an immersive CVE-based system for adults with neurodevelopment disorders (NDD) [86]. They developed four virtual tasks adapted from a well-known psychometric test used in clinical practice, the Wechsler Adult Intelligence Scale-IV or WAIS [87]. An equivalent physical experimental setup was created for one of the virtual scenarios as an experimental control. Twenty-four participants with NDD were divided into two groups: the CVE group (n=12; 6 pairs) and the physical interaction group (n=12; 6 pairs). The analysis included a comparison of conversational parameters, including total conversation time and the number of word utterances between the groups. The study reported a significantly higher number of words used in the CVE group compared to the physical interaction group in the same amount of time, which may indicate richer conversation quality for participants in the CVE group. As presented here, CVE-based systems can support complex social interactions where users can converse and interact in a natural setting, which may assist in successfully transitioning learned skills to the real world.

1.3.4 Intelligent Agents to support skill learning

The application of intelligent agents in serious games effectively facilitates users to learn and practice their skills in the games by providing users with prompts and directive responses [88]. An intelligent agent (IA) is designed using artificial intelligence (AI) technology to perceive the environment, such as human actions and behaviors, and then provide suitable human-like responses. There are several advantages of implementing IA in HCI systems. In a single-user serious game, an agent interacts with the human users while playing the games together. One role of an IA is, as an opponent that challenges the user to continue improving their skill. For instance, agents were embedded in board games such as chess and checkers [89]. These agents were designed to adaptively change the game patterns to continue challenging the human user, keeping them engaged while practicing their skills. In a serious collaborative game, an intelligent agent can take the role of a partner that can play the game with the user. The agent can communicate with the user while working together toward the same goal. For example, a cooperative puzzle game with an IA was designed to investigate cooperative skills in a human-agent interaction [90]. A human user partnered with an IA using turn-taking to solve the puzzle, where both of them have individual and combined goals. Another role that an agent can take is a strategic intelligent agent, where it is designed with the goal to benefit the most in the interaction; as such, the agent can either cooperate with the user or disagree with the user based on the goal of the game. Cuayáhuitl and colleagues designed an artificially intelligent agent that can play a strategic board game called Settlers of Catan [91]. In the board game, players can offer resources to other players and also reply to offers made by other players. Although an agent can serve as an interaction partner that can support social interaction, the interaction is still restrictive compared to human-human interaction. In a collaborative HCI system with human-human interaction, additional roles can be introduced for the agent, but research in this domain is still underexplored. In multi-user HCI systems such as CVE, the interactions mainly take place between two or more human users. An intelligent agent in such systems takes a supportive role rather than actively participating in the interaction. Isbister et al. designed a passive intelligent agent in a virtual social meeting application [92]. The agent monitors the interaction, understands the context of the conversation, provides suggestions on the conversation topic, and fades back into the background. The agent was not designed to participate in the interaction actively but was needed to maintain the interaction in the meeting, similar to a role of a moderator.

In both single-user and multi-user HCI systems with intelligent agents, one of the main challenges in the design of an intelligent agent is perceiving human behavior. By correctly perceiving a behavior, the agent will then provide a suitable individualized response to the user. In earlier systems with intelligent agents, perception of human behavior was simplified by using limited interaction modalities such as text-chat or selections of pre-defined actions. One of the early systems in this area, ELIZA, was designed by Weizenbaum in 1966 [93]. ELIZA could participate in natural language interaction with a human user by identifying keywords of a user-typed input sentence and then generating responses based on the keywords and predefined rules. In another example, Thinking Head, a VR-based system with a virtual tutor, aims to teach facial recognition and social situation understanding for children with ASD [94]. A predefined database was created to store individual responses based on expected inputs and generic responses for unexpected inputs. The virtual agent could generate appropriate speech and facial expressions based on the user's input, based on button clicks with pre-built texts. However, rule-based and predefined responses are not robust and flexible enough to perceive complex human interactions, specifically in a multi-user system. Human behavior is inherently stochastic, which means that for the same measurable action repeated over some time, the output measurement can be different every time for the same person, making it almost impossible to define individual input to represent an action.

One area of interest that researchers widely explore to interpret stochastic human data is pattern recognition and machine learning (ML). Pattern recognition allows for optimal predictions from a set of ambiguous data using probability and decision theories [95]. An intelligent agent using a probabilistic model can better predict human behavior and then provide flexible, accurate, and individualized responses to the user [96]. In [91], their study focused on applying a Deep Reinforcement Learning (DRL) method to train strategic conversation skills for the agent. The agent was trained to make offers that can give the highest pay-off in the long run by either: a) accepting an offer, b) rejecting an offer, or c) making a counteroffer. Results of the study indicated that the DRL method significantly outperformed several other methods, including random, rule-based, and supervised methods, in training the agent's conversational skills. Other than DRL, Hidden Markov Model (HMM) has become increasingly popular in perceiving human behavior such as speech recognition [97], spelling out words using hand gestures in American Sign Language (ASL) [98], dance movement recognition [99], and recognizing group actions [100]. Mihoub et al. presented a probabilistic modeling framework using Incremental Discrete Hidden Markov Model

(IDHMM) to model an intelligent agent capable of recognizing and generating multimodal joint actions in a face-to-face interaction [101]. The result indicated that IDHMM produced a much higher classification rate than a Support Vector Machines (SVM), with a mean cognitive state recognition rate of 92% compared to 81%, and a modest mean gaze generation rate of 49% compared to 43%. Zhang et al. implemented two layers of HMM to evaluate group actions in meetings [100]. The first HMM layer was used to recognize individual actions. Based on the outcome of this first layer, the second HMM layer then evaluated the group interaction actions. The model was tested with 59 public corpora of meeting data and results were compared to a single layer HMM. The result showed that the two layers HMM had a 70.3% accuracy while a single layer HMM only had 57.5% accuracy. Based on these studies, modeling an intelligent agent with probabilistic models can significantly improve the perception of human behavior, thus allowing the agent to provide better support to users in serious collaborative games.

1.3.5 Multimodal Data Analytics to assess skill learning

As early as the 1930s, psychologists have been collecting behavioral data to analyze human behaviors [102]. Advancements in the field of HCI research have provided researchers with access to an abundant amount of data, catalyzing a growing interest among researchers to gain insight into observable and underlying human behaviors from these data [103]. For example, studies have extracted data to either group user behavior [104], classify typical user profiles [105], or assess user performance [106]. Furthermore, advancements in sensors technology and peripheral devices have provided researchers access to a wide range of data from simple keypresses to involuntary human action, such as heart rate.

According to a meta-analysis review of 30 studies that compared affect detection accuracy between multimodal and unimodal data, researchers reported that multimodal accuracies were consistently better than unimodal accuracies [107]. This is further supported by the findings in a study by Mallol-Ragolta et al. where they reported best agreement score using a multimodal model compared to a unimodal model in a robotic empathy recognition system [108]. Driven by these promising results, recent HCI-related studies are using multimodal data capture to better understand human behaviors. For example, Bian et al. used multiple physiological features such as electrodermal activity and heart rate variability to represent affective states [109].

Although there are many studies that capture multimodal data in collaborative interactions, there are currently no standardized methods to measure teamwork and collaboration skills in group interactions. A study by Okada et al. used verbal and non-verbal measures to assess communication skills of individuals in different types of discussion tasks [110]. In the study, they captured data from speech and head movement information. Users' speech data were used to represent both verbal and non-verbal information. Researchers analyzed linguistic information such as parts of speech (PoS) and dialogue act (DA) using natural language processing (NLP) to get verbal information. They then extracted the length of speech, number of utterances, and the average length of utterance in a single session. Head movement data represented the non-verbal information. Based on the features extracted from the multimodal data, they tested eight regression models with different feature combinations and compared the results against human-coded evaluations of communication skills. They reported that the best regression model combined all features in both verbal and non-verbal interactions. Both studies found that certain quantitative measures can be analyzed to represent more than one feature. For example, speech data can be used to represent both verbal and non-verbal features in collaborative interactions [110] and dialogue content can provide social communication features (e.g., intention) and task performance features (e.g., topic/object) [103]. A comprehensive study by Echeverria et al. on multimodal analytics for group data serve as a reference in utilizing multimodal data in a group interaction [103]. The study created a collaboration matrix that connects a list of collaboration dimensions to a set of multimodal measures. Researchers applied the matrix for healthcare simulation scenarios to observe communication and teamwork dynamics between nurses while providing care to a patient. Using the matrix, researchers observed the dynamic communication patterns, physical localization and proximity, arousal level, and overall actions by each participant in the simulation. Results indicated that multimodal analytics could provide insight into group interactions and collaboration performance that could complement human observations. In line with the study conducted by [103], we conducted a literature survey to identify dimensions of collaboration for distributed task-oriented collaboration interaction. Meier et al. developed 9 dimensions of collaboration in a computer-supported video conference system between psychology and medical students [47]. The study was focused on problem-solving skills between individuals from different knowledge domains, as such the dimensions of collaboration in the study looked at different areas of speech such as joint information processing, motivation, and coordination. Parson conducted a study that looked at computer-supported collaboration interaction between children with autism and typically developing partners [113]. In the study, the researcher defined and evaluated both verbal and non-verbal behavior of the participants to identify interactional actions, peer-to-peer communication, and intervention needed. Using the evaluation scheme, the author successfully observed that autistic children showed less effective collaboration and actions compared to TD children, and they needed more prompts from the teacher to complete the tasks.

One of the core characteristics of individuals with ASD is a deficit in emotional reciprocity, where they have difficulties identifying and describing their own emotions [111]. They might show subtle changes in emotions or low manifestation of specific behavior that human observers may not capture but can be captured through multimodal measurements. One way multimodal analytics can benefit individuals with ASD is by analyzing the quantitative measures, allowing researchers to observe the hidden aspect of their behavior. For example, Zheng et al. created a multimodal HCI system to capture the occurrence of precursors to behavioral issues in individuals with ASD [112]. Clinicians and parents could anticipate the change in behavior and act sooner to stop the behavior before it happens.

1.4 Specific Aims and Summary of Research Work

My research focuses on designing and developing novel team-based activities in an intelligent collaborative virtual environment (CVE) to efficiently encourage, assess, and support teamwork and EF skills training. In this dissertation, the systems we developed were studied with individuals with Autism Spectrum Disorder (ASD) as an intervention tool for teamwork training. However, the system can be used by other individuals with learning disabilities and typically developing individuals. Chapter 2 to Chapter 7 discuss the studies in more details.

1.4.1 Specific Aim 1: Collaborative Games in CVE for Children with ASD

For this aim, we developed two CVE systems with various games to investigate the acceptance of the tasks among children with ASD (6 – 15 years old) and evaluate the system's feasibility to support collaborative behaviors and communication. Three main collaboration principles were applied to the collaborative games in both systems: a) turn-taking, b) joint action, and c) information sharing. The initial design and development of both systems were done by previous graduate students while I facilitated with the experimental procedures.

(a) <u>Collaborative puzzle games to measure social communication and collaboration skills of children with ASD</u>

In this study, we created a CVE system with a conversational agent capable of interacting and playing collaborative puzzle games with children with ASD. The objectives of the system were to evaluate the feasibility of the agent interacting with the children and evaluate the agent's accuracy to measure communication and collaboration skills in children with ASD based on speech and interaction data. Details of the system design and results from the study are discussed in Chapter 2.

(b) <u>Collaborative haptic games to train fine motor skills while supporting social communication and collaboration in children with ASD</u>

Next, we developed two collaborative haptic games that can support dyadic interaction between one child with ASD and one typically developing (TD) child to practice their collaboration and fine motor skills. Each child used a haptic device with an attached force gripper to control the movements of virtual objects in the CVE, where they can 'touch' and feel the 'weight' of the virtual objects they are manipulating. The objectives of this system were to observe changes in fine motor skills in children with ASD and how fine motor skills training influences social communication and collaboration between participants. Results indicated that children could communicate and work with each other while practicing their fine motor skills. Details of the system design and results are discussed in Chapter 3.

1.4.2 Specific Aim 2: Team-based Virtual Activities in CVE for Adults with ASD

Based on the knowledge we gained from developing the systems in Specific Aim 1 and the encouraging findings from the studies, we developed three new collaborative tasks set in a real-world employment setting for older individuals with ASD. We used the same collaborative principles for these tasks. Additionally, we also introduced principles of executive functioning skills such as working memory and time management within the tasks. We employed an inclusive design process by reviewing the designed tasks with various stakeholders, including industry experts, clinicians, certified trainers, and autistic individuals themselves. The tasks were updated based on their suggestions and feedbacks.

Our study evaluated the system's acceptability with adults with ASD, how well the tasks elicited teamwork behaviors and communication, and verified the multimodal data we captured. Results indicated preliminary acceptance of the CVE-based system, a positive impact of the collaborative tasks on supported teamwork skills practice for autistic and neurotypical individuals, and promising potential to quantitatively assess collaboration through multimodal data analysis. Details of the system and study results are available in Chapter 5.

1.4.3 Specific Aim 3: Quantitative Measures of Teamwork and Executive Functions through Accumulation of Multimodal Data in Dyadic Interactions

(a) Mapping of dimensions of teamwork and executive functions to multimodal data

As presented in Section 1.2.3, there has been significant interest in analyzing and interpreting multimodal data to represent complex skills such as teamwork and EF. However, multimodal analysis of autistic individuals' data is still limited. In our work, we captured multimodal data that include speech, eye gaze location, region of interests, controller button presses, physiological, and tasks related measures from individuals with ASD and their TD partner. Through an extensive literature survey, we defined specific dimensions of teamwork and EF within the collaborative tasks developed in Specific Aim 2. Then, we mapped these dimensions to the multimodal measures to get a quantitative evaluation of teamwork and EF skills. Details of the mapping dimensions of collaboration to the multimodal data are discussed in Chapter 5.

(b) Collaborative assessment task in CVE to evaluate quantitative measures of teamwork and executive functions

Next, we designed an additional collaborative task that was used as an assessment task to evaluate the performance of teamwork and executive functions of both participants with ASD and TD. The task was designed as a Lego-inspired block building task, where an individual with ASD worked with a TD peer to choose and build an animal figure using Lego blocks. A mapping of dimensions of collaboration to the corresponding multimodal data was also created. This task was validated and used as pre-test and post-test tasks in a study that explored the immediate effect of training with the virtual tasks developed in Specific Aim 2. The results demonstrate that the collaborative assessment task was able to capture the changes in the dimensions of collaboration through multimodal analytics. Chapter 6 describes in detail the design of the assessment task and how the dimensions of collaboration were incorporated within the task using multimodal data capture.

1.4.4 Specific Aim 4: Predictive Model of Participants' State in Collaborative Interaction for Effective Feedback Mechanism

(a) Using a rule-based model and a probabilistic model to predict user states in collaborative interaction

The collaborative puzzle game we developed in Specific Aim 1(a) has provided us with an understanding of using both rule-based and probabilistic models to discern human behavior by analyzing user's speech and controller input data. Using the same puzzle game, we wanted to explore real-time gaze data to perceive the user's gaze when interacting with a virtual avatar while playing a puzzle game to train joint attention skills in children with ASD. A simple rule-based model was developed to interpret the user's gaze data. Details of the study and results are discussed in Chapter 4. Although the model implemented in this system worked well as it was measuring only one skill, this rule-based model is not scalable and flexible enough for a more complex system such as our team-based CVE system. To address this limitation, we designed a prediction model that could recognize complex behavior in CVE-based interaction and domain beyond the puzzle game. We designed a Hidden Markov Model (HMM) to predict the teamwork behavior between individuals with ASD and their TD partners. We used the multimodal data captured in the study from Specific Aim 2 to train and design a reliable prediction model using HMM. We compared the HMM-predicted behavior to ground truth hand-labeled data and achieved an overall accuracy of 96.4%, whereas a rule-based prediction model for teamwork behavior had an overall accuracy of 77%. Details of the HMM design and development is available in Chapter 7.

(b) Real-time feedback mechanism to facilitate teamwork training

An effective training system or intervention tool can provide response while practicing the skills. Positive reinforcements can motivate users to continue with their actions, while a prompt can be helpful to point out the expected actions to the user. Responses can be provided either manually by a person observing the training session or embedded within the system through an intelligent agent. In our work, we used a finite state machine (FSM) to model the different types of feedback given by the intelligent agent. In the same study that we conducted in Specific Aim 3(b), we analyzed the results to observe the immediate effect of training with the collaborative tasks with embedded feedback mechanism in the experimental group and comparing the results to a control group. The results showed positive changes in collaboration and performance for autistic participants. A detailed discussion of the results is available in Chapter 6.

References

[1] Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. Autism, 20(4), 442-462.

[2] Trilling, B., & Fadel, C. (2009). 21st century skills: Learning for life in our times. John Wiley & Sons.

[3] 21st Century Skills and the Workplace, Microsoft, Pearson Report 2013

[4] Kulman, R., Slobuski, T., & Seitsinger, R. (2014). Teaching 21st century, executive-functioning, and creativity skills with popular video games and apps. Learning, Education and Games: Volume One: Curricular and Design Considerations, 1, 159.

[5] Slovák, P., & Fitzpatrick, G. (2015). Teaching and developing social and emotional skills with technology. ACM Transactions on Computer-Human Interaction (TOCHI), 22(4), 1-34.

[6] Salas, E., Burke, C. S., & Cannon?Bowers, J. A. (2000). Teamwork: emerging principles. International Journal of Management Reviews, 2(4), 339-356.

[7] Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. Human factors, 50(3), 540-547.

[8] Salas, E., Fowlkes, J. E., Stout, R. J., Milanovich, D. M., & Prince, C. (1999). Does CRM training improve teamwork skills in the cockpit?: Two evaluation studies. Human Factors, 41(2), 326-343.

[9] McEwan D., Ruissen G. R., Eys M. A, Zumbo B., & Beauchamp M. (2017). The Effectiveness of Teamwork Training on Te amwork Behaviors and Team Performance: A Systematic Review and Meta-Analysis of controlled Interventions. PLoS ONE, 12(1)

[10] Williams-Bell, F.M., Kapralos, B., Hogue, A. et al. Using Serious Games and Virtual Simulation for Training in the Fire Service: A Review. Fire Technol 51, 553-584 (2015)

[11] Paiva, P. V., Machado, L. S., Valença, A. M. G., Batista, T. V., & Moraes, R. M. (2018). SimCEC: a collaborative VR-based simulator for surgical teamwork education. Computers in Entertainment (CIE), 16(2), 1-26.

[12] Cai, P., Chandrasekaran, I., Cai, Y., Chen, Y., & Wu, X. (2017). Simulation-enabled vocational training for heavy crane operations. In Simulation and Serious Games for Education (pp. 47-59). Springer, Singapore.

[13] American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington: American Psychiatric Publishing, 2013

[14] Maenner MJ, Shaw KA, Bakian AV, et al. Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2018. MMWR Surveill Summ 2021;70(No. SS-11):1-16. DOI: http://dx.doi.org/10.15585/mmwr.ss7011a1external icon.

[15] Shattuck, PT. (2019, June 27), Growing numbers of young adults on the autism spectrum, Drexel University Life Course Outcomes, https://drexel.edu/autismoutcomes/blog/overview/2019/June/Growing-numbers-of-young-adults-on-the-autism-spectrum/

[16] Dietz PM, Rose CE, McArthur D, Maenner M. National and State Estimates of Adults with Autism Spectrum Disorder. Journal of Autism and Developmental Disorders. 2020

[17] Hendricks, D. (2010). Employment and adults with autism spectrum disorders: Challenges and strategies for success. Journal of vocational rehabilitation, 32(2), 125-134.

[18] Taylor, J. L., & Seltzer, M. M. (2011). Employment and post-secondary educational activities for young adults with autism spectrum disorders during the transition to adulthood. Journal of autism and developmental disorders, 41(5), 566-574.

[19] Baron-Cohen, S., Ashwin, E., Ashwin, C., Tavassoli, T., & Chakrabarti, B. (2009). Talent in autism: hyper-systemizing, hyper-attention to detail and sensory hypersensitivity. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1522), 1377-1383.

[20] Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. Journal of autism and developmental disorders, 36(1), 5-25.

[21] Kirchner, J. C., & Dziobek, I. (2014). Towards successful employment of adults with autism: A first analysis of special interests and factors deemed important for vocational performance. Scandinavian Journal of Child and Adolescent Psychiatry and Psychology, 2(2), 77-85.

[22] Waisman-Nitzan, M., Schreuer, N., & Gal, E. (2020). Person, environment, and occupation characteristics: What predicts work performance of employees with autism?. Research in Autism Spectrum Disorders, 78, 101643.

[23] Chen, W. (2012). Multitouch tabletop technology for people with autism spectrum disorder: A review of the literature. Procedia Computer Science, 14, 198-207.

[24] Bernard-Opitz V, Sriram N, Nakhoda-Sapuan S. Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. Journal of Autism and Developmental Disorders. 2001;31(4):377-384.

[25] Sung, C., Connor, A., Chen, J., Lin, C. C., Kuo, H. J., & Chun, J. (2019). Development, feasibility, and preliminary efficacy of an employment-related social skills intervention for young adults with high-functioning autism. Autism, 23(6), 1542-1553.

[26] Parsons, S., & Mitchell, P. (2002). The potential of virtual reality in social skills training for people with autistic spectrum disorders. Journal of intellectual disability research, 46(5), 430-443.

[27] Putnam, C., & Chong, L. (2008, October). Software and technologies designed for people with autism: what do users want?. In Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility (pp. 3-10).

[28] Almaguer, E., & Yasmin, S. (2019, July). A Haptic Virtual Kitchen for the Cognitive Empowerment of Children with Autism Spectrum Disorder. In International Conference on Human-Computer Interaction (pp. 137-142). Springer, Cham.

[29] Saiano, M., Pellegrino, L., Casadio, M., Summa, S., Garbarino, E., Rossi, V., ... & Sanguineti, V. (2015). Natural interfaces and virtual environments for the acquisition of street crossing and path following skills in adults with Autism Spectrum Disorders: a feasibility study. Journal of neuroengineering and rehabilitation, 12(1), 1-13.

[30] Bian, D., Wade, J. W., Zhang, L., Bekele, E., Swanson, A., Crittendon, J. A., ... & Sarkar, N. (2013, July). A novel virtual reality driving environment for autism intervention. In International Conference on Universal Access in Human-Computer Interaction (pp. 474-483). Springer, Berlin, Heidelberg

[31] Zheng, Z., Zhang, L., Bekele, E., Swanson, A., Crittendon, J. A., Warren, Z., & Sarkar, N. (2013, June). Impact of robot-mediated interaction system on joint attention skills for children with autism. In 2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR) (pp. 1-8). IEEE.

[32] Bekele, E., Zheng, Z., Swanson, A., Crittendon, J., Warren, Z., & Sarkar, N. (2013). Understanding how adolescents with autism respond to facial expressions in virtual reality environments. IEEE transactions on visualization and computer graphics, 19(4), 711-720.

[33] Powell, S. (1996, July). The use of computers in teaching people with autism. In Autism on the agenda: papers from a National Autistic Society Conference. London

[34] Bernard-Opitz, V., Ross, K., & Tuttas, M. L. (1990). Computer assisted instruction for autistic children. Annals of the Academy of Medicine, Singapore, 19(5), 611-616.

[35] Williams, C., Wright, B., Callaghan, G., & Coughlan, B. (2002). Do children with autism learn to read more readily by computer assisted instruction or traditional book methods? A pilot study. Autism, 6(1), 71-91

[36] Walsh, E., Holloway, J., & Lydon, H. (2018). An evaluation of a social skills intervention for adults with autism spectrum disorder and intellectual disabilities preparing for employment in Ireland: A pilot study. Journal of autism and developmental disorders, 48(5), 1727-1741.

[37] Agran, M., Hughes, C., Thoma, C. A., & Scott, L. A. (2016). Employment social skills: What skills are really valued?. Career Development and Transition for Exceptional Individuals, 39(2), 111-120.

[38] Schmutz, J. B., Meier, L. L., & Manser, T. (2019). "How effective is teamwork really? The relationship between teamwork and performance in healthcare teams: a systematic review and meta-analysis." BMJ open, 9(9), e028280. https://doi.org/10.1136/bmjopen-2018-028280

[39] Diamond, A. (2013). Executive functions. Annual review of psychology, 64, 135-168.

[40] Soft Skills to Pay the Bills, US Department of Labor, https://www.dol.gov/agencies/odep/program-areas/individuals/youth/transition/soft-skills

[41] Specialsterne: Assessment. https://www.specialisterneni.com/about-us/assessment/

[42] Neurodiversity hiring: Global diversity and inclusion at Microsoft. https://www.microsoft.com/en-us/diversity/inside-microsoft/cross-disability/neurodiversityhiring. (accessed Nov. 02, 2021).

[43] Praharaj, S., Scheffel, M., Drachsler, H., & Specht, M. (2018, September). Multimodal analytics for real-time feedback in co-located collaboration. In European Conference on Technology Enhanced Learning (pp. 187-201). Springer, Cham.

[44] Martinez-Maldonado, R., Gaševic, D., Echeverria, V., Fernandez Nieto, G., Swiecki, Z., & Buckingham Shum, S. (2021). What Do You Mean by Collaboration Analytics? A Conceptual Model. Journal of Learning Analytics, 8(1), 126-153.

[45] Palliya Guruge, C., Oviatt, S., Delir Haghighi, P., & Pritchard, E. (2021, October). Advances in multimodal behavioral analytics for early dementia diagnosis: A review. In Proceedings of the 2021 International Conference on Multimodal Interaction (pp. 328-340).

[46] Crescenzi Lanna, L. (2020). Multimodal Learning Analytics research with young children: A systematic review. British Journal of Educational Technology, 51(5), 1485-1504.

[47] A. Meier, H. Spada, and N. Rummel, "A rating scheme for assessing the quality of computer-supported collaboration processes," International Journal of Computer-Supported Collaborative Learning, vol. 2, no. 1, pp. 63-86, 2007, doi: 10.1007/s11412-006-9005-x

[48] Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. Computers & Education, 140, 103599.

[49] Herrera, F., Oh, S. Y., & Bailenson, J. N. (2020). Effect of behavioral realism on social interactions inside collaborative virtual environments. Presence, 27(2), 163-182.

[50] Riedl, M. O. (2019). Human?centered artificial intelligence and machine learning. Human Behavior and Emerging Technologies, 1(1), 33-36.

[51] Parsons, S., Mitchell, P., & Leonard, A. (2005). Do adolescents with autistic spectrum disorders adhere to social conventions in virtual environments? Autism, 9, 95-117

[52] Kelley, R., Tavakkoli, A., King, C., Nicolescu, M., Nicolescu, M., & Bebis, G. (2008, March). Understanding human intentions via hidden markov models in autonomous mobile robots. In Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction (pp. 367-374).

[53] Ali, M. R., Razavi, S. Z., Langevin, R., Al Mamun, A., Kane, B., Rawassizadeh, R., ... & Hoque, E. (2020, October). A virtual conversational agent for teens with autism spectrum disorder: Experimental results and design lessons. In Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents (pp. 1-8).

[54] Ali, M. R., & Hoque, E. (2017, September). Social skills training with virtual assistant and real-time feedback. In Proceedings of the 2017 ACM International Joint Conference on Pervasive and

Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers (pp. 325-329).

[55] Norouzi, N., Kim, K., Hochreiter, J., Lee, M., Daher, S., Bruder, G., & Welch, G. (2018, November). A systematic survey of 15 years of user studies published in the intelligent virtual agents conference. In Proceedings of the 18th international conference on intelligent virtual agents (pp. 17-22).

[56] STAN?IN, K., HOI?-BOŽI?, N., & SKO?I? MIHI?, S. (2020). Using Digital Game-Based Learning for Students with Intellectual Disabilities-A Systematic Literature Review. Informatics in education, 19(2), 323-341.

[57] Anastasiadis, T., Lampropoulos, G., & Siakas, K. (2018). Digital game-based learning and serious games in education. International Journal of Advances in Scientific Research and Engineering (ijasre), 4(12), 139-144.

[58] Sinha, G., Shahi, R., & Shankar, M. (2010, November). Human computer interaction. In 2010 3rd International Conference on Emerging Trends in Engineering and Technology (pp. 1-4). IEEE.

[59] Al Mahdi, Z., Naidu, V. R., & Kurian, P. (2019). Analyzing the Role of Human Computer Interaction Principles for E-Learning Solution Design. In Smart Technologies and Innovation for a Sustainable Future (pp. 41-44). Springer, Cham.

[60] Chandra, S., Sharma, G., Malhotra, S., Jha, D., & Mittal, A. P. (2015, December). Eye tracking based human computer interaction: Applications and their uses. In 2015 International Conference on Man and Machine Interfacing (MAMI) (pp. 1-5). IEEE.

[61] Colby, K. M. (1973). The rationale for computer-based treatment of language difficulties in nonspeaking autistic children. Journal of autism and childhood schizophrenia, 3(3), 254-260

[62] Delavarian, M., Bokharaeian, B., Towhidkhah, F., & Gharibzadeh, S. (2015). Computer-based working memory training in children with mild intellectual disability. Early Child Development and Care, 185(1), 66-74.

[63] Curatelli, F., Martinengo, C., Bellotti, F., & Berta, R. (2013, October). Paths for cognitive rehabilitation: from reality to educational software, to serious games, to reality again. In International Conference on Games and Learning Alliance (pp. 172-186). Springer, Cham.

[64] Fernández-Aranda, F., Jiménez-Murcia, S., Santamaría, J. J., Gunnard, K., Soto, A., Kalapanidas, E., ... & Penelo, E. (2012). Video games as a complementary therapy tool in mental disorders: PlayMancer, a European multicentre study. Journal of Mental Health, 21(4), 364-374.

[65] Salem, M., Kopp, S., Wachsmuth, I., Rohlfing, K., & Joublin, F. (2012). Generation and evaluation of communicative robot gesture. International Journal of Social Robotics, 4(2), 201-217.

[66] Schoene, D., Lord, S. R., Delbaere, K., Severino, C., Davies, T. A., & Smith, S. T. (2013). A randomized controlled pilot study of home-based step training in older people using videogame technology. PloS one, 8(3), e57734.

[67] Golomb, M. R., McDonald, B. C., Warden, S. J., Yonkman, J., Saykin, A. J., Shirley, B., ... & Burdea, G. C. (2010). In-home virtual reality videogame telerehabilitation in adolescents with hemiplegic cerebral palsy. Archives of physical medicine and rehabilitation, 91(1), 1-8.

[68] Bernardini, S., Porayska-Pomsta, K., & Smith, T. J. (2014). ECHOES: An intelligent serious game for fostering social communication in children with autism. Information Sciences, 264, 41-60.

[69] da Silva, C. A., Fernandes, A. R., & Grohmann, A. P. (2014, April). STAR: speech therapy with augmented reality for children with autism spectrum disorders. In International Conference on Enterprise Information Systems (pp. 379-396). Springer, Cham.

[70] Strickland, D., A virtual reality application with autistic children. Presence: Teleoperators & Virtual Environments, 1996. 5(3): p. 319-329.

[71] Charitos, D., et al. Employing virtual reality for aiding the organisation of autistic children behaviour in everyday tasks. in Proceedings of ICDVRAT. 2000.

[72] Matson, J.L., M.L. Matson, and T.T. Rivet, Social-skills treatments for children with autism spectrum disorders: An overview. Behavior modification, 2007. 31(5): p. 682-707.

[73] Engelbrecht, H., Lindeman, R. W., & Hoermann, S. (2019). A SWOT analysis of the field of virtual reality for firefighter training. Frontiers in Robotics and AI, 6, 101.

[74] Heldal, I., & Wijkmark, C. H. (2017, May). Simulations and Serious Games for Firefighter Training: Users' Perspective. In ISCRAM.

[75] Wang, P., Wu, P., Wang, J., Chi, H. L., & Wang, X. (2018). A critical review of the use of virtual reality in construction engineering education and training. International journal of environmental research and public health, 15(6), 1204.

[76] Zheng, Z. K., Sarkar, N., Swanson, A., Weitlauf, A., Warren, Z., & Sarkar, N. (2021). CheerBrush: A Novel Interactive Augmented Reality Coaching System for Toothbrushing Skills in Children with Autism Spectrum Disorder. ACM Transactions on Accessible Computing (TACCESS), 14(4), 1-20.

[77] Lee, K. (2012). The Future of Learning and Training in Augmented Reality. InSight: A Journal of Scholarly Teaching, 7, 31-42.

[78] Bao, H. J., Cheng, H. K., Vejayaratnam, N., Anathuri, A., Seksyen, S. K., Bangi, B. B., & Bakar, A. A. (2021). A STUDY ON HUMAN RESOURCE FUNCTION: RECRUITMENT, TRAINING AND DEVELOPMENT, PERFORMANCE APPRAISAL AND COMPENSATION. Journal of Global Business and Social Entrepreneurship (GBSE), 7(20).

[79] Jarrold, W., Mundy, P., Gwaltney, M., Bailenson, J., Hatt, N., McIntyre, N., ... & Swain, L. (2013). Social attention in a virtual public speaking task in higher functioning children with autism. Autism Research, 6(5), 393-410.

[80] L. Nguyen, D. Frauendorfer, M. Mast, and D. Gatica-Perez. Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior. IEEE Trans. on Multimedia, 16(4):1018-1031, 2014.

[81] Bozgeyikli, L., Bozgeyikli, E., Raij, A., Alqasemi, R., Katkoori, S., & Dubey, R. (2017). Vocational rehabilitation of individuals with autism spectrum disorder with virtual reality. ACM Transactions on Accessible Computing (TACCESS), 10(2), 1-25.

[82] Breland, J. S., & Shiratuddin, M. F. (2009). A Study on Collaborative Design in a Virtual Environment. International Journal of Learning, 16(3).

[83] Apostolellis, P., & Bowman, D. A. (2014, March). C-OLiVE: Group co-located interaction in VEs for contextual learning. In 2014 IEEE Virtual Reality (VR) (pp. 129-130). IEEE.

[84] Stichter, J. P., Laffey, J., Galyen, K., & Herzog, M. (2014). iSocial: Delivering the social competence intervention for adolescents (SCI-A) in a 3D virtual learning environment for youth with high functioning autism. Journal of autism and developmental disorders, 44(2), 417-430.

[85] Kandalaft, M. R., Didehbani, N., Krawczyk, D. C., Allen, T. T., & Chapman, S. B. (2013). Virtual reality social cognition training for young adults with high-functioning autism. Journal of autism and developmental disorders, 43(1), 34-44.

[86] Vona, F., Silleresi, S., Beccaluva, E., & Garzotto, F. (2020, November). Social matchup: Collaborative games in wearable virtual reality for persons with neurodevelopmental disorders. In Joint International Conference on Serious Games (pp. 49-65). Springer, Cham.

[87] Hartman, D. E. (2009). Wechsler Adult Intelligence Scale IV (WAIS IV): return of the gold standard. Applied neuropsychology, 16(1), 85-87.

[88] Frutos-Pascual, M., & Zapirain, B. G. (2015). Review of the use of AI techniques in serious games: Decision making and machine learning. IEEE Transactions on Computational Intelligence and AI in Games, 9(2), 133-152

[89] Schaeffer, J., Burch, N., Björnsson, Y., Kishimoto, A., Müller, M., Lake, R., ... & Sutphen, S. (2007). Checkers is solved. science, 317(5844), 1518-1522.

[90] Kulms, P., Mattar, N., & Kopp, S. (2015, August). An interaction game framework for the investigation of human-agent cooperation. In International Conference on Intelligent Virtual Agents (pp. 399-402). Springer, Cham.

[91] Cuayáhuitl, H., Keizer, S., & Lemon, O. (2015). Strategic dialogue management via deep reinforcement learning. arXiv preprint arXiv:1511.08099.

[92] Isbister, K., Nakanishi, H., Ishida, T., & Nass, C. (2000, April). Helper agent: Designing an assistant for human-human interaction in a virtual meeting space. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 57-64).

[93] Weizenbaum, J. (1966). ELIZA-a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36-45.

[94] Milne, M., Luerssen, M. H., Lewis, T. W., Leibbrandt, R. E., & Powers, D. M. (2010, July). Development of a virtual agent based social tutor for children with autism spectrum disorders. In the 2010 international joint conference on neural networks (IJCNN) (pp. 1-9). IEEE.

[95] Bishop, C. M. (2014). Bishop-Pattern Recognition and Machine Learning-Springer 2006. Antimicrob. Agents Chemother, 03728-14.

[96] Sosnowski, T., & Yordanova, K. (2020, June). A probabilistic conversational agent for intelligent tutoring systems. In Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments (pp. 1-7).

[97] Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257-286.

[98] Lu, S., Picone, J., & Kong, S. (2013). Fingerspelling alphabet recognition using a two-level hidden markov model. In Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp) (p. 1).

[99] McCormick, J., Vincs, K., Nahavandi, S., Creighton, D., & Hutchison, S. (2014, June). Teaching a digital performing agent: Artificial neural network and hidden markov model for recognising and performing dance movement. In Proceedings of the 2014 International Workshop on Movement and Computing (pp. 70-75).

[100] Zhang, D., Gatica-Perez, D., Bengio, S., & McCowan, I. (2006). Modeling individual and group actions in meetings with layered HMMs. IEEE Transactions on Multimedia, 8(3), 509-520.

[101] Mihoub, A., Bailly, G., & Wolf, C. (2013, October). Social behavior modeling based on incremental discrete hidden Markov models. In International Workshop on Human Behavior Understanding (pp. 172-183). Springer, Cham.

[102] Skinner, B. F. (1938). The behavior of organisms: An experimental analysis. D. Appleton-Century Company, incorporated.

[103] Echeverria, V., Martinez-Maldonado, R., & Buckingham Shum, S. (2019, May). Towards collaboration translucence: Giving meaning to multimodal group data. In Proceedings of the 2019 chi conference on human factors in computing systems (pp. 1-16).

[104] Dumais, S., Jeffries, R., Russell, D. M., Tang, D., & Teevan, J. (2014). Understanding user behavior through log data and analysis. In Ways of Knowing in HCI (pp. 349-372). Springer, New York, NY.

[105] Liu, Z., Wang, Y., Dontcheva, M., Hoffman, M., Walker, S., & Wilson, A. (2016). Patterns and sequences: Interactive exploration of clickstreams to understand common visitor paths. IEEE Transactions on Visualization and Computer Graphics, 23(1), 321-330.

[106] Wang, G., Zhang, X., Tang, S., Zheng, H., & Zhao, B. Y. (2016, May). Unsupervised clickstream clustering for user behavior analysis. In Proceedings of the 2016 CHI conference on human factors in computing systems (pp. 225-236).

[107] D'Mello, S., & Kory, J. (2012, October). Consistent but modest: a meta-analysis on unimodal and multimodal affect detection accuracies from 30 studies. In Proceedings of the 14th ACM international conference on Multimodal interaction (pp. 31-38).

[108] Mallol-Ragolta, A., Schmitt, M., Baird, A., Cummins, N., & Schuller, B. (2019, May). Performance analysis of unimodal and multimodal models in valence-based empathy recognition. In 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019) (pp. 1-5). IEEE.

[109] Bian, D., Wade, J., Swanson, A., Warren, Z., & Sarkar, N. (2015, February). Physiology-based affect recognition during driving in virtual environment for autism intervention. In International Conference on Physiological Computing Systems (Vol. 2, pp. 137-145). SCITEPRESS.

[110] Okada, S., Ohtake, Y., Nakano, Y. I., Hayashi, Y., Huang, H. H., Takase, Y., & Nitta, K. (2016, October). Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets. In Proceedings of the 18th ACM International Conference on Multimodal Interaction (pp. 169-176).

[111] Bird, G., & Cook, R. (2013). Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism. Translational psychiatry, 3(7), e285-e285.

[112] Zheng, Z. K., Staubitz, J. E., Weitlauf, A. S., Staubitz, J., Pollack, M., Shibley, L., ... & Sarkar, N. (2021). A Predictive Multimodal Framework to Alert Caregivers of Problem Behaviors for Children with ASD (PreMAC). Sensors, 21(2), 370.

CHAPTER 2: DESIGN OF AN INTELLIGENT AGENT TO MEASURE COLLABORATION AND VERBAL-COMMUNICATION SKILLS OF CHILDREN WITH AUTISM SPECTRUM DISORDER IN COLLABORATIVE PUZZLE GAMES

2.1 Abstract

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by core deficits in social interaction and communication. Collaborative puzzle games are interactive activities that can be played to foster the collaboration and verbal-communication skills of children with ASD. In this paper, we have designed an intelligent agent that can play collaborative puzzle games with children and verbally communicate with them as if it is another human player. Furthermore, this intelligent agent is also able to automatically measure children's task-performance and verbal-communication behaviors throughout game play. Two preliminary studies were conducted with children with ASD to evaluate the feasibility and performance of the intelligent agent. Results of Study I demonstrated the intelligent agent's ability to play games and communicate with children within the game-playing domain. Results of Study II indicated its potential to measure the communication and collaboration skills of human users.

2.2 Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by core deficits in social interaction and communication [1]. The estimated prevalence of ASD in the United States is 1 in 59, as reported by the Centers for Disease Control and Prevention [2]. The individual incremental lifetime cost associated with ASD is over $3.2 million [3]. With its high prevalence rate and associated costs, a wide range of studies have explored mechanisms to positively impact the social communications of children with ASD as well as the improvement of their long-term developmental outcomes [4], [5]. Although a cumulative literature review suggests that some interventions can have positive impacts on the lives of children with ASD and their families, many families struggle to access evidence-based care due to its high cost (often over $100/hour, with recommended intensity of at least 15 hours per week) and a shortage of trained clinicians [6], [7]. Therefore, an urgent need exists for inexpensive, accessible, and effective assistive therapeutic modalities for ASD intervention.

Computer-assisted interventions may offer an alternative intervention and assessment modality with reduced costs of care [8]. Many children with ASD have a natural affinity for computer-controlled environments [9] and exhibit a high level of engagement within these systems [10]. In addition, computer systems can provide controllable, replicable, and safe environments for children with ASD to practice social communication skills. As such, various kinds of computer-mediated intervention systems have been developed in order to understand and enhance the social communication skills of children with ASD [10],

[11]. Among these are collaborative game-based interventions, which usually target two users' ability to convey information to one another (communicate) and to work together to achieve a common goal (collaborate) [12].

Collaborative games poses several advantages relative to traditional intervention and assessment modalities. First, many children with ASD show a high level of engagement in computer-based collaborative games. Hourcade and colleagues designed four computer games that required two users to work together [13]. They analyzed users' collaborative interactions by manually coding the users' conversations within the system, and found that children with ASD spoke more sentences when they played these computer games as compared to non-computer games. Another advantage of collaborative computer games is that they can be designed to include strategies to elicit collaborative skills. Battocchi and colleagues designed collaborative puzzle games with an enforced collaboration rule, which required two users to take actions simultaneously, in or-der to encourage collaborations between the users [14]. They evaluated the effect of these games on users' collaborations by measuring users' task performance, such as task completion time and number of moved puzzle pieces. They found that these collaborative games, equipped with the enforced collaboration rule, have more positive effects on children with ASD, compared to games without such rules. Piper and colleagues de-signed and implemented a cooperative tabletop computer games for adolescents with Asperger's Syndrome [15]. They found that the cooperative computer games improved engagement, group work skills and confidence to interact in social activities within the population. Other previous literature on collaborative games with intervention strategies have also successfully investigated other collaborative behaviors of children with ASD, such as sharing [16], turn taking [17], and collaborative play [18].

In this work, we present the design of an intelligent agent able to play collaborative games with children with ASD, and simultaneously communicate with them during the games. In addition, the intelligent agent was designed to automatically measure collaboration and verbal-communication skills of children with ASD when they played these games. Such an intelligent agent may have the ability to 1) encourage collaborative interaction and communication in children with ASD; and 2) automatically evaluate the impacts of these collaborative games on these children. We also conducted two preliminary studies to evaluate the feasibility of the intelligent agent to interact with the target population, as well as its potential to measure their collaboration and verbal-communication skills.

The main challenge of designing such an intelligent agent is to understand the unrestricted human language using a computer program. Note that designing a computer program that can understand human language and conduct conversations as a human (i.e., Turing test) is yet to be solved from a technical point of view [19], [20]. Existing intelligent agents with conversation capabilities can only work in narrowly defined domains [19], [21], [22]. In our implementation, the intelligent agent was also designed with narrowly defined domains when communicating and playing games with children with ASD.

2.2.1 Intelligent Agents with Conversation Capabilities

Intelligent agents with conversation capabilities have been studied for several decades. One of the early systems in this area, ELIZA, was designed by Weizenbaum in 1966 [23]. ELIZA could make natural language conversation with human, by identifying keywords of a user-typed input sentence, and then generating responses based on the keywords and predefined rules. Since that time, similar methods have been widely applied to create chatbots to simulate intelligent conversation. One of the most powerful chatbots is A.L.I.C.E., which has the ability to engage in conversation using 40000 predefined rules [24]. This system, however, cannot provide information unless the required information has already been stored in the system. Chatbots and question-answering applications, such as Apple's Siri [25], are typically designed to answer general questions based on predefined question-answer pairs or online searching. They cannot be directly used for a specific domain, such as game playing, due to a lack of domain-specific knowledge.

The majority of existing intelligent agents with conversation capabilities were developed to conduct flexible conversations in narrowly defined domains, such as flight and travel booking [21], train information tracking [22], and for museum guides [19]. However, there is no common way that existing systems that have been developed [26], with variations in purpose, method of understanding linguistic meaning, complexity, robustness, and coverage of domains [27]-[29] to be a single system. Given that the goal of this work is to design an intelligent agent that can not only communicate but also play collaborative games, we review relevant works on intelligent agents with conversation capabilities for game playing.

2.2.2 Intelligent Agents with Conversation Capabilities for Game Playing

Many existing intelligent agents with conversation capabilities for game playing have been designed to assist humans in interactive games. One of the important applications in this area is the Non-Player Character (NPC) with conversational capability. For example, the adventure game, Zork-series [30], included NPCs that could parse and understand the words and phrases typed by players, and then show specific text-based information to assist the players in the game. Magerko and colleagues designed a game with NPCs that could take actions based on players' commands [31]. Although NPCs in these systems can support communication with players, the communication usually is less flexible, with a fixed-format. Generally, such fixed-format methods are not suitable for measuring flexible communication between users in collaborative games.

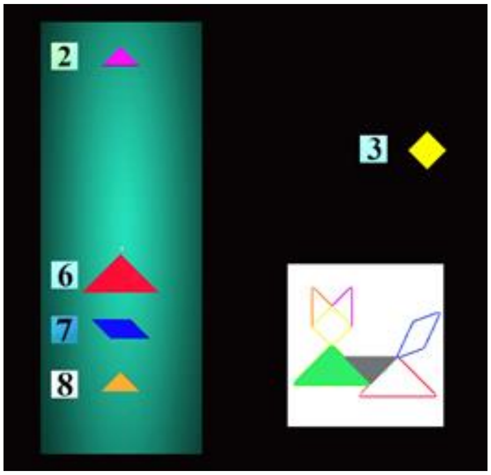


Figure 2-1: An example of the collaborative puzzle game

Only a few intelligent agents with conversation capabilities have been designed to support and measure flexible conversations within the collaborative game domain. Cuayáhuitl and colleagues designed an artificial intelligent agent that can play a strategic board game, called Settlers of Catan [32]. In the board game, players can offer resources to other players and they can also reply to offers made by other players. Their study focused on applying a Deep Reinforcement Learning (DRL) method to train conversational skills of the agent. Results of the study indicated that the DRL method significantly outperformed several other methods, including random, rule-based, and supervised methods, in training the agent's conversational skills. Kulms and colleagues designed an intelligent agent that could conduct text-based conversation as well as play a collaborative puzzle game [33]. In the collaborative puzzle game, the agent works together with a human to place differently shaped blocks in three steps:

1) one player, either the agent or the human, recommends one of two blocks to the other player;
2) the other player either accepts the recommendation and places the recommended block, or rejects the recommendation and chooses a different block;
3) the first player places the remaining block.

The two game actions, recommendation and acceptance/rejection, were used as measures of cooperation since they were indicative of competence, trust, and pursued goals. Unfortunately, to date very few results have been reported about the agent. These technologies provide important guidance about how to design intelligent agents to conduct conversations with a human and measure their communication behaviors. However, they were designed for a typically developed (TD) population, and could not be directly used for ASD intervention.

2.3 Collaborative Puzzle Games

In our previous study [34], we designed several computer-based collaborative puzzle games to encourage communication and collaborative interactions for children with ASD. Our collaborative puzzle games were designed based on tangram games [35], which require users to move seven puzzle pieces to form a specific shape. However, our collaborative games (see Fig. 2-1 for an example) were different from traditional tangram games since they required two users in different locations to interact with each other in order to play the game. In particular, these two users in two different locations played these games in a shared virtual environment and talk with each other through audio chat to exchange game information to complete the same tangram puzzle. An intelligent agent embedded into the virtual environment has the capability to interact with the user and also assess communication and collaboration skills of the user when they interact with each other and also with the agent. Details of the intelligent agent is described in the next section.

In order to encourage communicative and collaborative interactions, the collaborative puzzle games were equipped with two intervention strategies: 1) puzzle pieces could be moved together or individually; and 2) color of the puzzle pieces could be visible to one user or both users. In order to complete these games, both users needed to talk with each other to take turns moving the puzzle pieces, synchronize their actions to move the pieces together, or share color information. In the following sections, we first present the design and development of an intelligent agent that could talk and play these collaborative games with a child with ASD. Then, we discussed two case studies, which were conducted to test: 1) whether the intelligent agent could communicate and play the collaborative games with the children with ASD; and 2) whether the intelligent agent could generate features to measure the children's collaboration skills and verbal-communication skills.

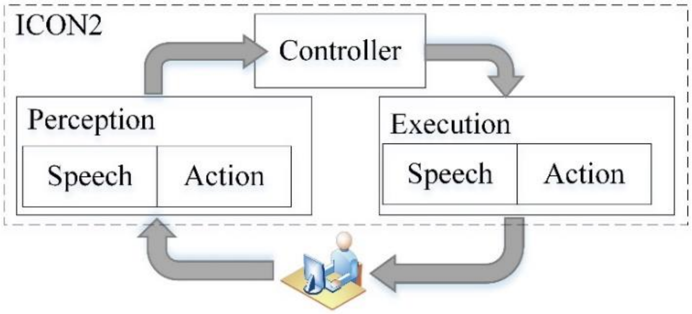


Figure 2-2: Overall system interaction diagram of ICON2

2.4 Intelligent Agent

2.4.1 Overall Description and Architecture

We designed an intelligent agent with the ability to elicit and assess COllaboratioN and COmmunicatioN (ICON2) skills of children with ASD through games and conversation tasks. The overall view of ICON2 is shown in Fig. 2-2. ICON2 can perceive a human's speech and game-related actions, i.e., what the human-partner says and what he/she does to play collaborative puzzle games. Then, it generates speech and game-related actions based on the perceived information. Finally, it executes these generated speech and game-related actions as responses to the human. ICON2 monitors input in real time without requiring the presence of a human therapist or coder. As ICON2 plays the games, as described below, it can assess collaborative and communicative aspects of the interaction through machine learning methods using several collaboration and communication features that were defined based on previous work [34].

A human user interacts with ICON2 when playing collaborative puzzle games. ICON2 acts as a virtual partner that is capable of conversing with the human user and also executes game actions during game play. ICON2 is aware of the game states, the rules of the games, and the layout of the virtual environment. As described in section 2.3, the puzzle games employ two configurations to promote collaboration, "turn-taking" and "move together." When the game is in the "turn-taking" configuration, the human user first moves a puzzle piece to the target image, and ICON2 observes the movement and waits for its own turn. If the human user does not make a move, ICON2 will prompt the user by saying, "This is a turn-taking game. It is your turn to move the puzzle piece." When it is ICON2's turn to move a puzzle piece, it asks the human user which piece it should move. It then "listens" to what the human user says and independently moves the identified piece to the target. When the game is in a "move together" configuration, ICON2 waits for the human user to communicate which piece to move together, verbally confirms the selection, and moves the piece together with the human user. If the human user does not verbally communicate which piece to move (e.g., User silently clicks on a piece to move), ICON2 will attempt to prompt user communication by asking, "Which piece should we move?"

ICON2 is aware of the human user game actions (puzzle piece locations and mouse clicks locations) and combines this information with verbal input from the human user. For example, a human user can opt to get color information from ICON2 in two ways:

1) Human user can ask ICON2, "What color is puzzle 5?" or
2) Human user can use the mouse to click on puzzle 5 and ask ICON2, "What is the color for this?"

ICON2 can correctly respond to both user actions with the color information of puzzle 5, even though in the later way the user did not specify the number of the puzzle piece.

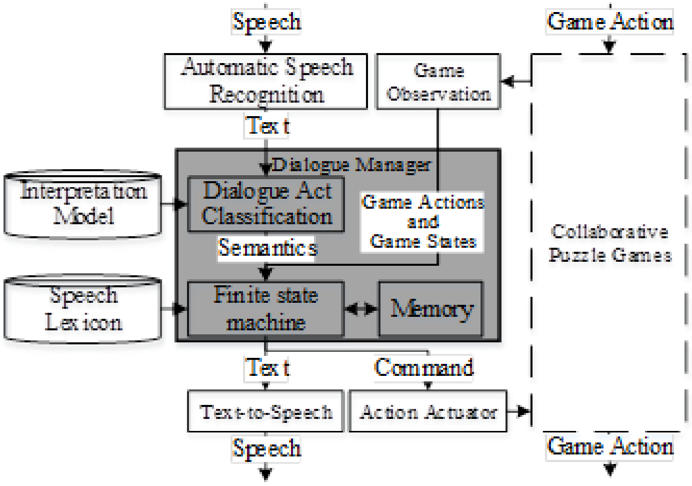


Figure 2-3: System architecture for ICON2

Different game characteristics are employed to encourage communication and collaboration in the human user and described in Table 2-2 and Table 2-3. For example, in both configurations, when the color information of the puzzle pieces is only available to one user (human user or ICON2), both partners have to exchange the color information in order to move the pieces to the target image. If prompted by the human user, ICON2 will always respond with the correct color information. When color information is not available for ICON2, it will ask the human user the color information before the puzzle piece can be moved to the target image.

ICON2 can provide assistance to human users when they fail to carry out necessary actions during game play. Since ICON2 shares the same goal of completing the puzzle games, it will move a puzzle piece if the human user does not after a certain period of time. After taking this independent action, ICON2 will re-iterate the current game rules to the human user by describing the game again. For example, “This is a turn-taking game. I have the color information. It is 22your turn to move a puzzle piece.” Or, “We need to move the puzzle piece together. Which piece do you want to move?”

The ICON2’s ability to communicate and play games was implemented with the architecture shown in Fig. 2-3. The architecture was composed of an Automatic Speech Recognition (ASR) module, a Game Observation (GO) module, a Dialogue Manager (DM) module, a Text-To-Speech (TTS) module, an Action Actuator (AA) module, and two databases, an Interpretation Model and a Speech Lexicon.

Each module of ICON2 was designed in order to support domain-related conversation and collaborative interactions in the puzzle games. The ASR module was used to perceive human’s speech inputs by transcribing the speech into text using Google Cloud Speech API for its low word error rate (approximately 8%). The GO module perceived game-related information such as human’s game actions and current game states. Information from these modules were then fed into the DM module, which was the main component of ICON2. The DM module was implemented using a hybrid method, which combines a dialogue act classifier and a finite state machine. The dialogue act classifier understood a human's domain-related speech using an interpretation model database as shown in Fig. 2-3. The finite state machine combined speech inputs, game-related inputs, and historical information to generate speech and game-related responses. All the historical information was stored in the memory of the DM module. In order to diversify the speech responses, the speech lexicon was used to map a speech semantic to different speech presentations. The TTS module used the Vuforia text recognition to transfer the text-based speech presentations to voice responses. The responses from the DM were then used in the TTS module and AA module to execute speech responses and to execute game-related actions respectively

Since the DM module was the core component of ICON2, we mainly focus on the design and development of the DM module.

TABLE 2-1

Dialogue Act Classes and Descriptions

Index	Name	Description	Example
1	Request color	Ask the color of a puzzle piece	What is the color of this puzzle piece?
2	Provide	Provide some information	It is red.
3	Direct movement	Direct ICON2 to move a puzzle piece	Move the green one.
4	Acknowledge	Acknowledge	Okay!
5	Request object	Ask about a puzzle piece	Which piece would you like to move? Which one is yellow?

2.4.2 Dialogue Manager

In our prior work, a total of 14 pairs of children (7 ASD/TD pairs and 7 TD/TD pairs) played collaborative puzzle games with each other [34]. The communication and game-playing behaviors of these users were analyzed, and then were used as the training data to design the communication and game-playing behaviors of ICON2.

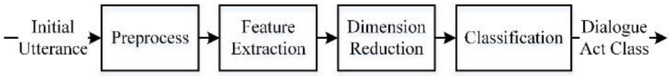


Figure 2-4: The process of the online classification

All the domain-related behaviors of these users can be presented as pairs of intentions and objects. An intention indicates the type of action a user plans to take in a collaborative puzzle game. Possible intentions in playing collaborative games include, to:

1) know the color of a puzzle piece;
2) drag a puzzle piece;
3) direct another user to drag a puzzle piece;

4) find a puzzle piece to move.

An object indicates a specific puzzle piece targeted by an intention. Possible values of the object can be any of the seven puzzle pieces or empty. In order to simulate the real users' collaboration and verbal-communication behaviors, ICON2 must be able to:

1) understand a human's intention;
2) find a targeted object; and
3) generate appropriate speech and game-related responses.

Besides the communication and game-playing behaviors, ICON2 should also be able to evaluate users' communication skills. Therefore, the development of its core component, i.e., the DM module, must conform to the following requirements:

1) ICON2 needs to both communicate and play games. Therefore, the DM module must have the ability to combine both speech and game-related inputs, and generate both speech and game-related responses.
2) ICON2 is required to assess a user's collaboration and verbal-communication skills. Therefore, the DM module must be able to gather or generate relevant features for these assessments.
3) As a partner to play collaborative games, ICON2 must be able to act proactively in a dialogue, i.e., to take initiative, rather than being purely responsive. This means that the DM module must not only respond to user's speech but also initiate a conversation.

In order to fulfill these requirements, we developed the communication and game-playing behaviors of ICON2 in three steps: i) understanding human spoken natural language and collecting game-related inputs, ii) detecting intention and object from the speech and game-related inputs, and iii) generating speech and game-related responses. These three steps together could enable both communication and game-playing capabilities. The speech and game-related inputs gathered in the first step were not only important for ICON2 to communicate and play games but also useful for ICON2 to evaluate skills of its human-partner. In the third step, some rules were included so that ICON2 can also initiate conversations in addition to responding to the users. In what follows, we describe the first step in section 2.4.2.1 and section 2.4.2.2, the second step in section 2.4.2.3, and the third step in section 2.4.2.4.

2.4.2.1 Language understanding

In order to be understandable for a computer, human language is typically represented using a set of messages, where each set has a finite number of messages and each message is associated with a particular action [36]. One way to represent human utterances is using combinations of dialogue acts and slots. A dialogue act is the specialized performative function that an utterance plays in language [37]. A slot is a variable that stores specific domain-related information of human's utterances [38]. Using a combination of dialogue act and slot to represent an utterance has been shown to be useful [39]-[41]. For example, AT&T's spoken dialogue system may represent a caller's request as "Report (payment)," where "report" is the dialogue act and "payment" is the slot [42]. We used combinations of dialogue acts and slots to represent utterances because dialogue acts were shown to be useful in evaluating the collaboration and verbal-communication behaviors in collaborative learning environments [43].

Dialogue acts and slots are usually domain-dependent. We defined dialogue acts and slots in our game playing domain based on the conversations recorded in our previous study. Five classes of dialogue acts (request color, provide, direct movement, acknowledge, and request object) are included in the game playing domain. The descriptions of these dialogue act classes are shown in Table 2-1. We then defined seven slots (color, id, object, action, policy, subject, and out-of-domain) and several slot words for each slot to represent users' utterances. The slot words of the first six slots could describe specific features of the collaborative puzzle games. For example, the color slot words, i.e., red, green, yellow, blue, pink, orange, and gray, described the colors of all the puzzle pieces in the games. The out-of-domain slot words (such as name, food, school, weekend, and Facebook), which were extracted from the out-of-domain utterances in the previous study, were used to describe out-of-domain information.

The dialogue act class of each utterance was computed using an interpretation model, while the slot words of each utterance were extracted by mapping each word of the utterance with all predefined slot words. We built an interpretation model using the recorded conversations in our previous study, and utilized the model to recognize a dialogue act of each utterance that a user used to communicate with ICON2. The interpretation model for this research was a Support Vector Machine with Radial Basis Function (SVM-RBF) kernel. The model was built using 136 data samples collected from our previous human-human interactions study using the process as shown in Fig. 2-4. First, we replaced each recognized slot word with its slot type since all the words belonging to a slot perform similar functionality in conducting utterances. This preprocessing procedure can reduce feature dimension. Second, we extracted multiple syntactic and word sequence features, including unigrams, bigrams, part of speech, and dependency types. It has been found that unigrams and bigrams are the most useful word sequence features in dialogue act classification [44], [45]. Parts of speech and dependency types are also useful structure features in dialogue act classification [46]. Natural Language Toolkit [47] was used for the feature extraction. After the feature extraction, we reduced the dimension of the features using Principal Component Analysis (PCA). Finally, the low-dimensional features together with labels of these training data samples were input to train the SVM-RBF model. A 5-fold cross validation was used to select hyper parameters of the SVM-RBF model. The feature extraction method, the PCA model, and the SVM-RBF model were also used for on-line classification.

2.4.2.2 Game-related inputs

The game-related inputs, including the human-partner game actions and current game states, are gathered from the collaborative games. Some examples of human-partner game actions are:

1) No action for a certain time-duration.
2) Dragging a puzzle piece.
3) Clicking on a puzzle piece.
4) Stop dragging a puzzle piece.

The values of these variables are extracted from the human-partner actions within the games.

The current game states are used to represent the interactive environment. They are composed of multiple parameters, such as the color of each puzzle piece, the position of each puzzle piece, and the target position. Two important parameters for current game state are the 1) color visibility; and 2) piece translation control, which are used to determine the features of each game, as mentioned in section 2.3. These game-related inputs are meaningful for ICON2 to detect intention, detect object, and generate responses.

2.4.2.3 Intention and object detection

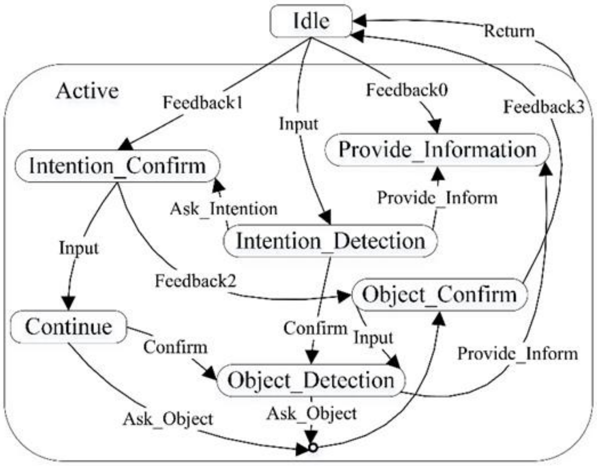


Figure 2-5: Finite state machine in the Dialogue Manager Module

A Finite State Machine (FSM), shown in Fig. 2-5, was developed to combine the speech and game-related inputs and generate speech and game-related responses. In general, spoken dialogue systems that are capable of speech and non-speech interactions can be implemented using two methods: rule-based and data-driven methods. The rule-based methods update information and generate responses using predefined rules [48]. Expertise is required to define these rules [49]. The data-driven methods, such as reinforcement learning [50], can generate models automatically from training data. However, gathering enough training data is challenging in most cases [51]. This work used a FSM with some predefined rules to combine inputs and generate outputs given the limited training data. In particular, ICON2 detects the intentions and objects by combining human-partner's speech and game-related inputs, and generates responses based on the detected intention and object using the FSM as shown in Fig. 2-5. The interaction detection and object detection are implemented using an "Intention_Detection" state and an "Object_Detection" state in the FSM. If the information is incomplete, the FSM also includes the "Intention_Confirm" and "Object_Confirm" states in order for ICON2 to 1) clarify unclear information; and 2) gather lost information. Responses are generated based on the detected intention and object, and are provided to the human-partner in the "Provide_Information" state.

The logic for intention detection can be summarized using the structure shown in Fig. 2-6. The tree-structure simplifies the intention-detection procedure by dividing it into multiple steps. In the first step, ICON2 detects out-of-domain utterances based on a rule: if an utterance has out-of-domain slot words, the utterance is an out-of-domain utterance. As mentioned in the introduction section, existing spoken dialogue systems are usually designed to operate over a limited and definite domain [52]. To ensure satisfactory user experience, sthe spoken dialogue system must be able to detect out-of-domain (OOD) utterances, and provide feedback to the user when OOD utterances are detected. Previous literature has applied classification methods to explicitly model OOD utterances for OOD detection [53]. However, collecting enough training data to model OOD utterances is time-consuming and laborious. Given the limited training data samples of this study, it is hard to create an OOD model with acceptable accuracy. Therefore, we used a rule-based method to detect OOD utterances in the current study. This method has been proven to be useful in the system, as discussed in the results session. Other advanced OOD detection methods will be explored in our system in the future.

The tree-structure and embedded rules also enable ICON2 to handle ambiguity in natural language. Ambiguity in natural language means that an utterance has multiple meanings. ICON2 can reduce language ambiguity using associated game-related inputs and dialogue history based on rules. For example, if a user says "red," she/he may intend to provide either the color information or to direct ICON2 to move the red puzzle piece. If the current game state indicates that color is visible to ICON2 or the dialogue history includes a request for a puzzle piece to move, the user intends to direct ICON2 to move the red puzzle piece.

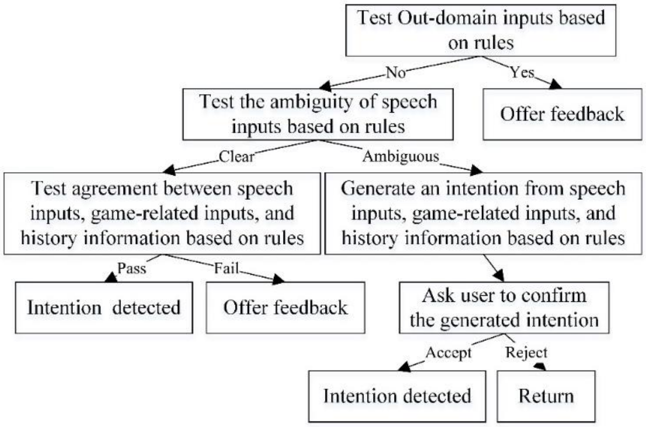

Test Out-domain inputs based on rules
No
Yes
Test the ambiguity of speech inputs based on rules
Offer feedback
Clear
Ambiguous
Test agreement between speech inputs, game-related inputs, and history information based on rules
Generate an intention from speech inputs, game-related inputs, and history information based on rules
Pass
Fail
Intention detected
Offer feedback
Ask user to confirm the generated intention
Accept
Reject
Intention detected
Return

Figure 2-6: The logic for intention detection

A weighted average method as shown in (1) uses both speech and game-related information in order to detect which puzzle piece is the targeted object. Equation (1) calculates the similarity between a puzzle piece and the targeted object. A targeted object is usually described using multiple characteristics, such as color of the object, index of the object, and actions on the object. In (1), different characteristics are presented using different terms, such as T_{color} , T_{index} , and $T_{position}$. The value of each term can be 1 (if the term matches the user's inputs) or 0 (if the term does not match the user's inputs). Each characteristic has a weight, such as W_{color} , W_{index} , and $W_{position}$ to reflect how important this characteristic is in the object detection. The values of these weights are predefined based on domain knowledge. ICON2 calculated a similarity value for each object based on (1). The object with the highest value is the targeted object. This method has the advantage to handle complex information in dialogue when describing an object.

$$W_{total} = W_{color} \times T_{color} + \cdots + W_{index} \times T_{index} \qquad (1)$$

2.4.2.4 Response generation

Based on the detected intention, detected object, and dialogue history, the DM module generates speech and game-related responses based on a set of carefully designed IF-THEN rules. For example, IF the intention is out-of-domain, THEN the agent provides feedbacks that reflects the out-of-domain nature of the conversation, such as "Hey! I only know something about the game we are playing. Let's play the game!"

ICON2 cannot only respond to human conversation but also initiate conversations. The ability to initiate a conversation enables ICON2 to act proactively in a dialogue, i.e., to take over the initiative, rather than being purely responsive. This leads to a more natural conversation and may facilitate communication in children with ASD within the system. The capability to initiate conversations is implemented using feedback events (“Feedback0,” “Feedback1,” “Feedback2,” and “Feedback3”) in the FSM (Fig. 2-5). These feedback events are triggered by a human’s game actions and are used to initiate an appropriate conversation. For example, if the human-partner has no action for 10 seconds, the “Feedback0” event is triggered and ICON2 may ask “I can see all the colors. Just ask when you need any.” if the color is invisible to the human.

ICON2 may say different sentences to express the same idea. A speech lexicon was used in order to generate different expressions. The speech lexicon stores multiple expressions for each idea. In real time, ICON2 can randomly select one of the expressions as the speech response. For example, if ICON2 wants to ask color of a puzzle piece, it may say:

1) “What is the color?”
2) “Could you tell me the color?”
3) “Is it red or green?”

A sample dialogue is shown in Fig. 2-7

Agent: We need to move pieces together during this game, I have all the colors.

Human: What is the color of this one?
(Human clicks on a puzzle piece)

Agent: That one is red.

Agent: Let’s move the red one together.
(Agent and human starts moving the puzzle piece)

(Human has no action for a while)

Agent: Which puzzle piece do you want to move?

Human: Number six.

Agent: It is a yellow one. Move number six. (Agent starts moving the puzzle piece)

Figure 2-7: A sample dialogue (All game actions are shown in parentheses)

2.5 Preliminary Studies

We conducted two preliminary studies to test i) whether ICON2 could interact and communicate with children with ASD to play collaborative puzzle games; and ii) whether ICON2 could generate meaningful features to measure the communication and collaboration skills of the children.

2.5.1 Collaborative Puzzle Games

In Study I, we used seven collaborative puzzle games developed in our previous study [34]. The variation in these games was implemented by manipulating two game features: 1) who can move the puzzle pieces; and 2) who can see the colors of the puzzle pieces. The characteristics of the seven collaborative puzzle games are shown in Table 2-2. Take Game_11 for example, where both users could see all the colors of the puzzle pieces, and they needed to take turns moving the pieces one by one. And for Game_15, only the human user could move the puzzle pieces, but ICON2 had the color information for the puzzle pieces. This forced the human user to ask ICON2 for the color information before choosing the puzzle piece to move to the target.

TABLE 2-2

Characteristics of Collaborative Puzzle Games in Preliminary Study I

Game name	Who can move puzzle pieces	Who can see color of puzzle pieces
Game_11	Users take turns	Both users
Game_12	Users take turns	Both users
Game_13	Both users together	Both users
Game_14	ICON2	Human user
Game_15	Human user	ICON2
Game_16	Both users together	Both users
Game_17	Both users together	Both users

In Study II, nine collaborative puzzle games were designed to elicit communication and collaboration of users. In these games, the puzzle pieces could be moved by taking turns, one at a time, or moved together. The colors of these puzzle pieces were visible to only one of the users or both users. As a result, the human user needed to talk with ICON2 to synchronize their actions, or to share color information. The characteristics of these games are shown in Table 2-3. Take Game_29 for example: in this game, only one user could see the colors of puzzle pieces, but both users needed to drag puzzle pieces together to a moving target area. Therefore, both users were required to converse with each other to share color information as well as to synchronize their actions in this game.

2.5.2 Participants and Experimental Procedure

Across both studies, a total of 10 children with ASD (5 children in each study) were recruited to interact with ICON2. Participants were recruited through an existing university-based clinical research registry. All participants had clinical diagnoses of ASD from a licensed psychological provider, had IQ scores higher than 70, and were capable of using phrase speech. To obtain current levels of autism symptomatology, parents completed the Social Responsiveness Scale, Second Edition (SRS-2) [54] and Social Communication Questionnaire Lifetime Total Score (SCQ) [55]. All study procedures were approved by the Vanderbilt University Institutional Review Board (IRB) with associated procedures for informed assent and consent.

TABLE 2-3

Characteristics of Collaborative Puzzle Games in Preliminary Study II

Game Name	Who can move puzzle pieces	Who can see color of puzzle pieces	Whether the target is moving
Game_21	Users take turns	Both users	No
Game_22	Users take turns	Only one user	No
Game_23	Users take turns	Only one user	No
Game_24	Both users together	Both users	No
Game_25	Both users together	Only one user	No
Game_26	Both users together	Only one user	No
Game_27	Both users together	Both users	Yes
Game_28	Both users together	Only one user	Yes
Game_29	Both users together	Only one user	Yes

Study I

The goal of this preliminary study was to test whether ICON2 could play games and communicate with children with ASD in the collaborative puzzle game domain. Five children with ASD participated in this study, and their characteristics are shown in Table 2-4. Each of the participants completed a one-visit experiment that lasted approximately 30 minutes. At the beginning of the experiment, the participant was shown both audio and text introduction about how to play the collaborative games with ICON2. Then the participant played seven collaborative puzzle games, as previously mentioned in Table 2-2. Finally, the participants completed a paper survey consisting of 6 items assessing user feedback regarding their interactions with ICON2. As seen in Table 2-7, each item consisted of a Likert scale with 1 being the most negative and 5 being the most positive. Research assistants explained the instructions to the participants and answered any questions that arose.

TABLE 2-4

The Characteristics of the Five Participants in Study I

Age	Gender	SRS-2 total raw score	SCQ current total score
Mean (SD)	Female/male	Mean (SD)	Mean (SD)
10.42 (3.31)	2/3	99.20 (21.65)	16.80 (5.36)

Study II

This preliminary study was aimed at testing whether the intelligent agent had the potential to generate meaningful features to measure communication and collaboration skills of children with ASD. Five children with ASD, different from the Study I participants, took part in this study (characteristics shown in Table 2-5). Each participant completed a one-visit experimental session. At the very beginning of the experiment, participants were shown an introduction explaining how to play games in the collaborative virtual environment (CVE). Then the participants played nine collaborative puzzle games in a random order.

Two researchers watched video recordings of the experiments and rated the communication and collaboration skills of the participants in order to provide the ground truth of these skills. They rated all the participants' skills on a binary rating scale with a value 1 or 0 after each game within a session (total of 9 games per participants). Values of the binary rating scale indicated whether the human raters felt the participants had a high level (value 1) or a low level (value 0) of communication and collaboration skills, respectively. These two human raters utilized the same rating scheme and rated the videos independently. The inter-rater agreement of the binary ratings was analyzed using a Cohen's Kappa method, which is a commonly used method to measure inter-rater agreement for categorical items [56]. The inter-rater agreement of the binary rating was 87.15%.

TABLE 2-5

The Characteristics of the Five Participants in Study II

Age	Gender	SRS-2 total raw score	SCQ current total score
Mean (SD)	Female/male	Mean (SD)	Mean (SD)
13.91 (1.91)	1/4	100 (14.40)	20.25 (7.41)

2.6 Skill Measurements Procedure

We now present the procedure to measure both communication and collaboration skills. The system generated task-performance and verbal-communication features to represent the participants' behaviors when they interacted with ICON2 in the CVE. Then we applied machine learning methods to measure these skills based on the system-generated features.

2.6.1 System-Generated Features

The system automatically generated multiple verbal-communication and task-performance features, which were designed based on previous literature in the field. All the features and their definitions are shown in Table 2-6. Previous literature demonstrated that dialogue act features, such as requests for information [27], providing information [28], and acknowledging other people's actions [29], were useful in understanding group discussion behaviors of both children with ASD and TD children. In addition, word frequency and sentence frequency have been found useful to reflect the behaviors of children with ASD during collaborative puzzle games [26]. Bauminger-Zviely and colleagues found that the success frequency and failure frequency reflected important aspects of collaborative behaviors of children with ASD in collaborative puzzle games [30]. White and colleagues reported that the dragging time and collaboration time features could reflect collaborative efficiency of children with ASD when they played collaborative puzzle games with their TD peers [31]. In our system, all the features shown in Table 2-6 were generated by the system in real-time and recorded for offline analysis.

2.6.2 Skill Measurements

TABLE 2-6

Automatically Generated Verbal-Communication and Task-Performance Features

Index	Feature	Description
1	Word frequency	How many words a user speaks per minute
2	Request color frequency	How many times per minute a user asks color information
3	Provide frequency	How many times per minute a user provides game information
4	Direct movement frequency	How many times per minute a user directs movements
5	Acknowledge frequency	How many utterances belong to acknowledgements
6	Request object frequency	How many times per minute a user asks for objects
7	Sentence frequency	How many utterances a user speaks in a minute
8	Success frequency	How many puzzle pieces have been successfully moved to the target area
9	Failure frequency	How many times a user fails in moving puzzle pieces

10	Collaboration time	The time duration of puzzle pieces being moved by two users simultaneously in a minute
11	Dragging time	The total time duration of a user dragging puzzle pieces
12	Collaborative movement ratio	The ratio of collaboration time and dragging time

We built machine learning models to measure participants' communication and collaboration skills using the system-generated features. In particular, we trained machine learning models to classify a data sample, which included all system-generated features of a game, into a binary-class, i.e., a high level of skills or a low level of skills. First, we applied Principal Component Analysis (PCA) to reduce the feature dimension. Then we trained a Support Vector Machine with Radial Basis Function (SVM-RBF) model to measure communication skills using the system-generated features and ratings of communication skills on a binary scale, and trained another SVM-RBF model to measure collaboration skills using the features and rating of the collaboration skills on a binary scale. We selected SVM-RBF kernel as the machine learning method for the classification because this method usually performs well in classifying data with a small sample size [57]. The performance of these models in measuring these skills was evaluated using their classification accuracies, which were computed using a 6-fold cross-valuation method.

2.7 Results

Overall, ICON2 worked as designed in this study. All participants completed their experiments. Unfortunately, experimental data of one participant in Study I was lost because the system crashed during the game for unknown reasons.

The data from Study I were analyzed to evaluate whether ICON2 could play the collaborative games and communicate within the game-playing domain with the participants. Data from Study II were analyzed to determine whether ICON2 had the potential to measure both communication and collaboration skills of the participants.

2.7.1 Results of Study I

Data for Study I include the survey scores provided by the participants after the experiment, the speech data, and the game action information collected during the collaborative game play.

Results of the distributed survey, as shown in Table 2-7, indicated that children with ASD enjoyed communicating and interacting with ICON2. The participants felt comfortable talking with ICON2 with an average score of 4 on a 1-5 Likert scale, where 1 meant very uncomfortable and 5 meant very comfortable. They could be understood by ICON2 with an average score of 3.8/5 and could also understand ICON2, with an average score of 4.2/5. They reported that it was easy to play the game with ICON2, as indicated by an average score 4.4/5 on Question 5, where 1 meant very difficult and 5 meant very easy. In addition, they enjoyed playing the games with ICON2 with an average score of 4.4/5, where 1 meant dislike very much and 5 meant like very much.

TABLE 2-7

Survey Results

Index	Questions	Mean	Standard deviation
1	Do you feel comfortable talking with ICON2 [1 very uncomfortable, 2 uncomfortable, 3 neutral, 4 comfortable; 5 very comfortable]	4	1
2	Do you feel ICON2 can understand you very well [1 strongly disagree; 2 disagree; 3 neutral; 4 agree; 5 strongly agree]	3.8	0.84
3	Do you feel you can understand ICON2 very well [1 strongly disagree; 2 disagree; 3 neutral; 4 agree; 5 strongly agree]	4.2	0.45
4	How quickly did ICON2 respond to you [1 very slowly; 2 slowly; 3 neutral; 4 quickly; 5 very quickly]	4.4	0.55
5	Overall, how easy do you think it is to play the game with ICON2 [1 very difficult; 2 difficult; 3 neutral; 4 easy; 5 very easy]	4.4	0.89
6	Overall, how much do you like to play the games with ICON2 [1 dislike very much; 2 dislike; 3 neutral; 4 like; 5 like very much]	4.4	0.55

A total of 249 utterances were spoken by the participants and a total of 374 utterances were generated by ICON2 in Study I. All utterances spoken by participants were found to be in-domain utterances and labeled as such, and no utterance was found to be out-of-domain. This result was in line with our previous human-human interaction study [34], in which children with ASD used very few (<0.01%) out-of-domain utterances when playing these games with their TD peers.

We asked a human coder to label these input utterances offline using the five dialogue act classes that were defined in Table 2-1. These manually labeled utterances were used as the ground truth for the classification. Results of dialogue act classification are shown in Table 2-8. The accuracy of the dialogue act classification of ICON2 was 67.47%, which was much higher than the random accuracy of 20%. These accuracies were computed based on the 249 utterances. Request Object feature had very low utterance frequency to calculate the classification accuracy, as such the value were set to 0 and omitted from further analysis. Please note that the classification accuracy was computed in real time, and the test data were independent of the training data.

Human coding results indicated that ICON2 had the potential to appropriately initiate conversations as well as to reply to the participants' speech. ICON2 generated two kinds of utterances:

1) "initiation," which was an utterance used to initiate a conversation;
2) "reply," which was an utterance used to reply to an initiated conversation by a participant.

We defined that all the utterances generated by the feedback events of the FSM were "initiations," and all the other utterances were "replies". In this study, ICON2 generated 161 initiations and 190 replies. 82.93% of the 161 initiations were labeled as appropriate initiations; while 89.33% of the 190 replies were labeled as appropriate replies. These results indicate that ICON2 demonstrates potential to communicate with the participants with ASD when they play puzzle games. Note that the accuracy of appropriate replies (89.33%) is much higher than the accuracy of dialogue act classification (67.47%), which suggests that ICON2 could reply appropriately even when it misunderstood a human's language. This was because ICON2 could reduce language ambiguity by combining the language with game-related inputs, as discussed in section 2.4.2.2.

TABLE 2-8

Dialogue Act Classification Results in Study I

		Target class					
		Request Color	Provide	Direct Movement	Acknowledge	Request Object	Sum
Classification results	Request Color	6.02%	0.40%	0.80%	0	0	7.23%
	Provide	0	37.35%	0.80%	0.80%	0	38.96%
	Direct Movement	0	5.62%	18.47%	2.81%	0	26.91%
	Acknowledge	0	20.08%	0.40%	5.62%	0	26.10%
	Request Object	0	0	0.80%	0	0	0.80%
	Sum	6.02%	63.45%	21.29%	9.24%	0	100.00%

We then used the game action and game states data to generate collaborative movement ratio. This is the ratio of the time duration when both human user and ICON2 simultaneously move a puzzle piece to the time duration when an individual user drags the piece. Results of collaborative movement ratio indicated that ICON2 could play collaborative games with these children. The average collaborative movement ratio of children with ASD when interacting with ICON2 in this study was 0.10, which was comparable to the ratio of 0.11 when children with ASD interacted with their TD peers in our previous study [58]. The collaborative movement ratio was a meaningful feature to measure collaborative efficiency when children with ASD played these collaborative puzzle games [34]. This result may indicate that ICON2 could effectively collaborate with children with ASD in the context of the games.

2.7.2 Results of Study II

In Study II, we wanted to test whether the ICON2 system could accurately generate verbal-communication features. Similar to Study I, we first generated the ground truth for these features using a human coding methodology. A human rater watched videos recorded during the experiments, manually transcribed the participants' speech to text, and labeled each sentence with one of the five predefined dialogue acts. The labels were used as the ground truth of the features. The accuracy of the five-class dialogue act classification was 69.10%, which was much higher than the random accuracy, 20%, of a five-class classification. Detailed results of the dialogue act classification are shown in Table 2-9. These accuracies were computed based on 1332 spoken sentences from the participants. The speech recognition errors and the dialogue act classification errors together led to errors of the system-generated verbal-communication features.

TABLE 2-9

Dialogue Act Classification Accuracies in Study II

		Target class					
		Request Color	Provide	Direct Movement	Acknowledge	Request Object	Sum
Classification results	Request Color	0.60%	0.07%	0.07%	0	0	0.74%
	Provide	0.07%	47.49%	5.76%	3.74%	0	57.06%
	Direct Movement	0	18.18%	17.47%	0.75%	0	36.40%
	Acknowledge	0	0.45%	0.60%	4.71%	0	5.76%
	Request Object	0	0.07%	0	0	0	0.07%
	Sum	0.67%	66.26%	23.90	9.20%	0	100%

An error rate of a feature is the ratio of the value difference between a system-generated feature and its true feature to the value of the true feature. The calculated error rate of each verbal-communication feature is shown in Table 2-10. The sentence frequency feature had the lowest error rate (0.0566). This result indicates that the system has the potential to accurately generate the sentence frequency feature. However, the utterance frequency of features Request Color and Request Object were very low resulting in very high error rate values. We removed these features from the analysis. We also present the ratio of the number of sentences in each dialogue act to the total number of sentences. This ratio was useful to understand the error rate of the corresponding verbal-communication feature.

TABLE 2-10

Error Rate of Each System-Generated Feature

System-generated Feature	Error rate	Ratio of the number of sentences in a dialogue act class to the total number of sentences
Word frequency	0.1289	--
Request color frequency	1.0000	0.0055
Provide frequency	0.3527	0.5027
Direct movement frequency	0.6408	0.4611
Acknowledge frequency	0.5789	0.0266
Request object frequency	1.0000	0.0041
Sentence frequency	0.0566	--

Table 2-11 shows the high accuracies the system managed to achieve in measuring the verbal-communication skills and collaboration skills using the system-generated features above. Since each game generated a data sample, we had 45 data samples for measurements (9 games, 5 participants). The accuracy to assess the binary communication skills based on these features was 89.68% using the SVM-RBF model discussed in section 2.6.2. This accuracy was computed with 30 data samples belonging to the high level of communication skills, while 15 data samples belonged to the low level of communication skills. The collaboration skills were assessed with a 75.40% accuracy with 28 data samples belonging to the high level of collaboration skills and 27 data samples belonging to the low level of collaboration skills. In addition, we present accuracies to measure these binary skills with balanced data samples. The balanced data were generated by randomly under-sampling the majority class, which is a commonly used resampling technique to improve classification performance in unbalanced datasets [59].

TABLE 2-11

Accuracy to Assess Both Communication and Collaboration Skills Based on the System-Generated Features

Index	Which skills to measure?	Data sample size (high level /low level)	Accuracy	Accuracy of balanced data
1	Communication skills	30/15	89.68%	79.20%
2	Collaboration skills	28/17	75.40%	74.95%

2.8 Conclusions

In this paper, we designed an intelligent agent that could communicate and play games with children with ASD in a CVE as well as generate meaningful features to measure their communication and collaboration skills. Results of the two preliminary studies presented here indicate the potential of ICON2 to 1) communicate and collaborate with children with ASD in the CVE as indicated by the self-report results; and 2) generate meaningful features to measure communication and collaboration skills of the participants as indicated by high accuracies of these measurements.

In particular, we found that ICON2 could appropriately initiate conversations and respond to the participants' conversation in Study I. ICON2 generated 82.93% appropriate initiations and 89.33% appropriate replies when interacting with the children with ASD. These accuracies are comparable to results of other intelligent agents with conversation capabilities designed for TD individuals [19], [60], [61]. Given differences in data sample numbers and task domains, it is hard to directly compare numerical results of different systems in this area. However, we believe that the communication capability of ICON2 are comparable to existing systems by comparing the numerical results available in the literature. For example, Kopp and colleagues designed a conversational agent as a museum guide to communicate with museum visitors. The agent could understand visitors' utterances by mapping keywords with 138 rules. The agent could correctly respond to visitors' 50423 utterances with an accuracy of 63% [19]. Tewari and colleagues designed a question-answer system to help improve reading skills of children in the lowest socio-economic status [60]. The system could correctly answer questions with an accuracy of 86%, which was computed with 346 utterances. However, this system could not initiate conversations and did not support non-speech interactions. Ramin and colleagues designed a spoken system to assist elderly users about their weekly planning. The system could respond to elderly users with 84.8% accuracy, which was computed from only 46 utterances [61].

ICON2 has the potential to evaluate communication and collaboration skills of the participants as seen in Study II. The system could accurately generate verbal-communication features as indicated by the low error rates of these features. For example, the sentence frequency feature had a low error rate 0.0566. All the features together could measure these skills with high accuracies using machine learning models. The accuracy to measure the communication skills was 89.58%, while the accuracy to measure the collaboration skills was 75.40%. Although these machine learning models were built offline, they could be used for real-time measurements in the future. The results indicate that the system has the potential to automatically measure both communication and collaboration skills in human-agent interactions based on these system-generated features. Automated systems for capturing, labeling, and measuring communicative overtures could, in the future, augment our ability to systematically measure change in important social-communication therapy goals. This has the potential to reduce costs associated with human observation and coding as well as reducing subjective bias in behavioral observation. That being said, in the current work the system required intensive human-coder classification in order to develop and optimize our models. Future, use of such a paradigm will ultimately have to overcome this system development cost to move toward larger scale use.

The errors that occurred when the system generated verbal-communication features were because of errors in speech recognition and errors in dialogue act classification. Errors of the word frequency and the sentence frequency features were due to errors of the speech recognition; while errors of other verbal-communication features, such as the Request Color frequency, Provide frequency, and Direct Movement frequency, were due to both the speech recognition errors and the dialogue act classification errors, as shown in Table 2-10. This might be the reason why the word frequency and sentence frequency features had the lowest error rates. We also found a high error rate for the Request Object frequency feature. This may be because the participants spoke only a few Request Object sentences, as indicated by the small ratio of the Request Object frequency to the sentence frequency in Table 2-10. As a result, a few incorrectly detected Request Object sentences could lead to a high error rate. We found similar results regarding the Request Color frequency feature.

While we have presented a novel hybrid method to develop this intelligent agent for meaningful measurements within the tangram puzzles domain with varying configurations (colors, no colors, turn taking, move together), ICON2's communication behaviors could be extended to other domains by modifying the hybrid method. ICON2's generated speech responses within the game-playing domain based on some rules can be extended to other domains by modifying these rules. After adjusting the variables of the hybrid method, ICON2 will be able to communicate and interact with users in other domains. Also, ICON2 design was not adaptive, where the system performed at the same level for users from different age and developmental group. It would be worth exploring the influence of varying the type of game as well as incorporating different difficulty levels to the communication and collaboration skills of the participants.

Although the present work is promising, readers are advised to exercise caution in interpreting the results more generally due to several limitations of the current work. First, the sample size was small, and the experimental design consisted of only one session. Please note that the goal of the present study was to design an intelligent agent that could play collaborative games and communicate within the game-playing domain to automatically measure important aspects of interactions in a CVE with preliminary studies. Results of the preliminary studies indicated that this intelligent agent has the potential to interact with children with ASD as well as automatically generate meaningful features to measure both communication and collaboration skills of children with ASD. In the next step, we will utilize this system for real-time measurements with more participants and with a longer study duration.

Second, the use of binary scale (0 = low, 1 = high) to rate the communication and collaboration skills may not be sufficient to provide in depth and continuous measure of these skills. The binary scale was used as an initial step to assess the feasibility of the agent without adding complexity of the analysis. Moving forward, work beyond proof of concept could possibly explore a more refined rating scheme that would be able to provide in depth rating of both skills.

Third, the training data used to build the SVM-RBF model for the dialogue act classification was relatively small. While the accuracy (67.47%) of the classifier in Study I and the accuracy (69.10%) of the classifier in Study II were much higher than the random accuracy (i.e., 20%) of a five-class classifier, more training data may yield a classification model with higher accuracy. In addition, the out-of-domain detection method in this paper was limited. Future studies should aim to develop more efficient methods for out-of-domain detection.

Fourth, the system-generated features were limited as well. We only explored 12 features for the measurements in the current study. Human behaviors, such as eye gaze, body language, and facial expression, could also provide important information in peer-mediated interactions. However, features to represent these behaviors have not been explored in this study. In the future, these features will be captured using eye gaze recognition, gesture recognition, and emotion recognition in order to understand the non-verbal communications.

And unlike an actual human partner, ICON2 has the potential to crash or fail. This could cause user frustration. As mentioned earlier in the section, the system did crash and caused data loss for one session, but the system was recovered right away to not cause further disruption to the experiment Despite these limitations, the performance of the games and interactions of the participants with their partners and the system itself were not affected and further contributes to the collaborative learning literature by proposing a novel way to automatically measure communication and collaboration skills of children with ASD within a CVE using an intelligent agent. Results of the two studies indicated that the presented intelligent agent was tolerated and apparently engaging and enjoyable to the participants, as well as demonstrate its potential to automatically measure important aspects of interactions in a CVE. The scope of the current work was to design the intelligent agent and preliminarily assess its capability to capture both communication and collaboration skills of children with ASD when they interacted with the intelligent agent in a CVE. This is a necessary first step before intelligent agents could be strategically deployed to assess these skills during peer-to-peer interactions within similar collaborative environments.

References

[1] C. G. M. Snoek, M. Worring, and A. W. M. Smeulders, “Early versus late fusion in semantic video analysis,” in *Proc. 13th ACM Int. Conf. Multimedia, MM 2005*, 2005, pp. 399–402, doi: 10.1145/1101149.1101236.

[2] J. Baio et al., “Prevalence of Autism Spectrum Disorder among children aged 8 years-Autism and Developmental Disabilities Monitoring Network,” 2014, MMWR Surveillance Summaries. doi: http://dx.doi.org/10.15585/mmwr.ss6706a1.

[3] G. Peacock, D. Amendah, L. Ouyang, and S. D. Grosse, “Autism Spectrum Disorders and Health Care Expenditures,” *J. Dev. Behav. Pediatr.*, vol. 33, no. 1, pp. 2–8, Jan. 2012, doi: 10.1097/DBP.0b013e31823969de.

[4] M. L. Sundberg and J. W. Partington, *Teaching Language to Children with Autism Or Other Developmental Disabilities*, Pleasant Hill, California, USA: Behavior Analysts, Inc., 1998.

[5] N. Bauminger, “The facilitation of social-emotional understanding and social interaction in high-functioning children with autism: Intervention Outcomes,” *J. Autism Dev. Disord.*, vol. 32, no. 4, pp. 283–298, Aug. 2002, doi: 10.1023/A:1016378718278.

[6] S. J. Rogers, “Empirically supported comprehensive treatments for young children with autism,” *J. Clin. Child Psychol.*, vol. 27, no. 2, pp. 168–179, 1998, doi: 10.1207/s15374424jccp2702_4.

[7] H. Cohen, M. Amerine-Dickens, and T. Smith, “Early intensive behavioral treatment: Replication of the UCLA model in a community setting,” *J. Dev. Behav. Pediatr.*, vol. 27, pp. S145-S155, 2006. doi: 10.1097/00004703-200604002-0001.

[8] R. C. Pennington, “Computer-assisted instruction for teaching academic skills to students with Autism Spectrum Disorders: a review of literature,” *Focus Autism Other Dev. Disabl.*, vol. 25, no. 4, pp. 239–248, Dec. 2010, doi: 10.1177/1088357610378291.

[9] D. Moore, Yufang Cheng, P. Mcgrath, and N. J. Powell, “Collaborative virtual environment technology for people with autism,” *Focus Autism Other Dev. Disabl.*, vol. 20, no. 4, pp. 231–243, 2005, doi: 10.1177/10883576050200040501.

[10] U. Lahiri, E. Bekele, E. Dohrmann, Z. Warren, and N. Sarkar, “A physiologically informed virtual reality based social communication system for individuals with autism,” *J. Autism Dev. Disord*, vol. 45, pp. 919-931, 2015, doi: 10.1007/s10803-014-2240-5.

[11] V. Bernard-Opitz, N. Sriram, and S. Nakhoda-Sapuan, “Enhancing social problem solving in children with Autism and normal children through computer-assisted instruction,” *J. Autism Dev. Disord.*, vol. 31, no. 4, pp. 377–384, Aug. 2001, doi: 10.1023/A:1010660502130.

[12] H. Noor, F. Shahbodin, and N. C. Pee, “Serious game for autism children: review of literature,” *World Academy Sci., Eng. and Technol. Int. J. Psychol. and Behav. Sci.*, vol. 6, pp. 554-559, 2012, doi: doi.org/10.5281/zenodo.1333272.

[13] J. P. Hourcade, S. R. Williams, E. A. Miller, K. E. Huebner, and L. J. Liang, “Evaluation of tablet apps to encourage social interaction in children with Autism Spectrum Disorders,” in *Conf. Human Factors Comput. Syst – Proc.*, 2013, pp. 3197–3206, doi: 10.1145/2470654.2466438.

[14] A. Battocchi, F. Pianesi, D. Tomasini, M. Zancanaro, G. Esposito, P. Venuti, A. Ben Sasson, E. Gal, and P. L. Weiss, “Collaborative puzzle game: A tabletop interactive game for fostering collaboration in children with Autism Spectrum Disorders (ASD),” in *Proc. ACM Int. Conf. Interactive Tabletops and Surfaces*, 2009, pp. 197–204, doi: 10.1145/1731903.1731940.

[15] A. M. Piper, E. O’Brien, M. R. Morris, and T. Winograd, “SIDES: A cooperative tabletop computer game for social skills development,” in *Proc. ACM Conf.Comput. Supported Cooperative Work, CSCW*, 2006, pp. 1–10, doi: 10.1145/1180875.1180877.

[16] D. D. Curtis and M. J. Lawson, "Exploring collaborative online learning," *J. Asynchronous Learn. Netw.*, vol. 5, pp. 21-34, 2001, doi: 10.24059/olj.v5i1.1885.

[17] M. Zancanaro, F. Pianesi, O. Stock, P. Venuti, A. Cappelletti, G. Iandolo, M. Prete, and F. Rossi, "Children in the museum: an environment for collaborative storytelling," in *PEACH-Intell. Interfaces Museum Visits*, 2007, pp. 165-184. doi: 10.1007/3-540-68755-6_8.

[18] 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, doi: https://doi.org/10.1177/1362361312451526.

[19] S. Kopp, L. Gesellensetter, N. C. Krämer, and I. Wachsmuth, "A conversational agent as museum guide-design and evaluation of a real-world application," in *Int. Workshop Intell. Virtual Agents*, 2005, pp. 329-343, doi: 10.1007/11550617_28.

[20] J. Cauell, T. Bickmore, L. Campbell, and H. Vilhjálmsson, "Designing embodied conversational agents," *Embodied Conversational Agents*, Cambridge, Massachusetts, USA: MIT Press, 2000, pp. 29-63.

[21] B. Pellom, W. Ward, J. Hansen, R. Cole, K. Hacioglu, J. Zhang, X. Yu, and S. Pradhan, "University of Colorado dialog systems for travel and navigation," in *Proc.1st Int. Conf. Human Lang. Technol. Res.*, 2001, pp. 1-6, doi: 10.3115/1072133.1072225.

[22] H. Aust, M. Oerder, F. Seide, and V. Steinbiss, "The Philips automatic train timetable information system," *Speech Commun.*, vol. 17, pp. 249-262, 1995, doi: https://doi.org/10.1016/0167-6393(95)00028-M.

[23] J. Weizenbaum, "ELIZA-a computer program for the study of natural language communication between man and machine," *Commun. ACM*, vol. 9, pp. 36-45, 1966, doi: https://doi.org/10.1145/365153.365168.

[24] A. Shawar and E. Atwell, "A chatbot system as a tool to animate a corpus," *ICAME J.: Int.Comput. Archive Modern and Medieval English J.*, vol. 29, pp. 5-24, 2005, doi: 10.1.1.110.9786.

[25] J. Aron, "How innovative is Apple's new voice assistant, Siri?," *New Scientist*, vol. 212, p. 24, 2011, doi: 10.1016/S0262-4079(11)62647-X.

[26] J. F. Allen, D. K. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent, "Toward conversational human-computer interaction," *AI Mag.*, vol. 22, p. 27, 2001, doi: https://doi.org/10.1609/aimag.v22i4.1590.

[27] J. Glass, "Challenges for spoken dialogue systems," in *Proc.1999 IEEE ASRU Workshop*, 1999, doi: 10.1.1.30.5112.

[28] M. F. McTear, "Spoken dialogue technology: enabling the conversational user interface," *ACM Comput. Surveys (CSUR)*, vol. 34, pp. 90-169, 2002, doi: 10.1145/505282.505285.

[29] M. Eskenazi, "An overview of spoken language technology for education," *Speech Commun.*, vol. 51, pp. 832-844, 2009, doi: 10.1016/j.specom.2009.04.005.

[30] J. Brusk and T. Lager, "Developing natural language enabled games in (Extended) SCXML," in *Proc. Int. Symp. Intell. Techniques Comput. Games and Simul.* (Pre-GAMEON-ASIA and Pre-ASTEC), Shiga, Japan, March, 2007, pp. 1-3.

[31] B. Magerko, J. Laird, M. Assanie, A. Kerfoot, and D. Stokes, "AI characters and directors for interactive computer games," in *Proc. 16th Conf. Innov. Appl. AI*, San Jose, California, USA, July 2004, pp. 877-883. 2004, doi: 10.5555/1597321.1597339.

[32] H. Cuayáhuitl, S. Keizer, and O. Lemon, "Strategic dialogue management via deep reinforcement learning," 2015. [Online]. Available: arXiv:1511.08099. doi: 10.17861/6c6de69e-25ea-4836-b443-44b312354fac.

[33] P. Kulms, N. Mattar, and S. Kopp, "An interaction game framework for the investigation of human-agent cooperation," in *Int. Conf. Intell. Virtual Agents*, 2015, pp. 399-402, doi: 10.1007/978-3-319-21996-7_43.

[34] L. Zhang, M. Gabriel-King, Z. Armento, M. Baer, Q. Fu, H. Zhao, A. Swanson, M. Sarkar, Z. Warren, and N. Sarkar, "Design of a mobile collaborative virtual environment for autism intervention," in *Int. Conf. Universal Access in Human-Computer Interaction*, 2016, pp. 265-275, doi: 10.1007/978-3-319-40238-3_26.

[35] M. S. Kanbar, "Tangram game assembly," U.S. Patent 4298200, Nov. 3, 1981.

[36] B-H. Juang and S. Furui, "Automatic recognition and understanding of spoken language-a first step toward natural human-machine communication," *Proc. IEEE*, vol. 88, no. 8, pp. 1142-1165, Aug. 2000, doi: 10.1109/5.880077.

[37] A. Stolcke, N. Coccaro, R. Bates, P. Taylor, C. Van Ess-Dykema, K. Ries, E. Shriberg, D. Jurafsky, R. Martin, and M. Meteer, "Dialogue act modeling for automatic tagging and recognition of conversational speech," *Comput. Linguist.*, vol. 26, pp. 339-373, 2000, doi: 10.1162/089120100561737.

[38] J. D. Williams and S. Young, "Partially observable Markov decision processes for spoken dialog systems," *Comput. Speech and Lang.*, vol. 21, pp. 393-422, 2007, doi: https://doi.org/10.1016/j.csl.2006.06.008.

[39] T-H. Wen, M. Gasic, D. Kim, N. Mrksic, P.-H. Su, D. Vandyke, and S. Young, "Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking," in *Proc. 16th Annu. Meeting Special Interest Group Discourse and Dialogue*, Prague, Czech Republic, Sept. 2015, pp. 275-284, doi: 10.18653/v1/W15-4639.

[40] P. Tsiakoulis, C. Breslin, M. Gasic, M. Henderson, D. Kim, M. Szummer, B. Thomson, and S. Young, "Dialogue context sensitive HMM-based speech synthesis," in *2014 IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, Florence, Italy, pp. 2554-2558, doi: 10.1109/ICASSP.2014.6854061.

[41] S. Zhu, L. Chen, K. Sun, D. Zheng, and K. Yu, "Semantic parser enhancement for dialogue domain extension with little data," in *2014 IEEE Spoken Lang. Technol. Workshop (SLT)*, South Lake Tahoe, Nevada, USA, pp. 336-341, doi: 10.1109/SLT.2014.7078597.

[42] N. Gupta, G. Tur, D. Hakkani-Tur, S. Bangalore, G. Riccardi and M. Gilbert, "The AT&T spoken language understanding system," in *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 14, no. 1, pp. 213-222, Jan. 2006, doi: 10.1109/TSA.2005.854085.

[43] A. Soller, A. Martínez, P. Jermann, and M. Muehlenbrock, "From mirroring to guiding: A review of state of the art technology for supporting collaborative learning," *Int. J. Artif. Intell. Ed.*, vol. 15, pp. 261-290, Dec 2005, doi: 10.5555/1434935.1434937.

[44] J. Fürnkranz, "A study using n-gram features for text categorization," *Austrian Res. Inst. Artificial Intelligence*, vol. 3, pp. 1-10, 1998, doi: 10.1.1.49.133.

[45] K. Samuel, S. Carberry, and K. Vijay-Shanker, "Dialogue act tagging with transformation-based learning," in *Proc. 17th Int. Conf. Comput. Linguistics*, Montreal, Quebec, Canada, 1998, vol. 2, pp. 1150-1156, doi: 10.3115/980691.980757.

[46] K. E. Boyer, E. Y. Ha, R. Phillips, M. D. Wallis, M. A. Vouk, and J. C. Lester, "Dialogue act modeling in a complex task-oriented domain," in *Proc.11th Annu. Meeting Special Interest Group Discourse and Dialogue*, Tokyo, Japan, 2010, pp. 297-305, doi: 10.5555/1944506.1944561.

[47] S. Bird, "NLTK: the natural language toolkit," in *Proc. ACL-02 Workshop Effective Tools and Methodologies Teaching Natural Lang. Process. and Comput. Linguistics*, Philadelphia, Pennsylvania, USA, 2006, vol. 1, pp. 69-72, doi: 10.3115/1118108.1118117.

[48] S. Larsson and D. R. Traum, "Information state and dialogue management in the TRINDI dialogue move engine toolkit," *Nat. Lang. Eng.*, vol. 6, pp. 323-340, Sept. 2000, doi: 10.1017/S1351324900002539.

[49] D. DeVault, A. Leuski, and K. Sagae, "Toward learning and evaluation of dialogue policies with text examples," in *Proc. SIGDIAL 2011 Conf.*, pp. 39-48, doi: 10.5555/2132890.2132896.

[50] J. D. Williams and S. Young, "Scaling POMDPs for spoken dialog management," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 15, no. 7, pp. 2116-2129, Sept. 2007, doi: 10.1109/TASL.2007.902050.

[51] T. Paek and R. Pieraccini, "Automating spoken dialogue management design using machine learning: An industry perspective," *Speech Commun.*, vol. 50, pp. 716-729, Aug. 2008, doi: 10.1016/j.specom.2008.03.010.

[52] I. Lane, T. Kawahara, T. Matsui, and S. Nakamura, "Out-of-domain utterance detection using classification confidences of multiple topics," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 15, no. 1, pp. 150-161, Jan. 2007, doi: 10.1109/TASL.2006.876727.

[53] P. J. Durston, M. Farrell, D. Attwater, J. Allen, H-K. J. Kuo, M. Afify, E. Fosler-Lussier, and C-H. Lee, "OASIS natural language call steering trial," in *7th European Conf. Speech Commun. and Technol.*, 2001, doi: 10.1.1.127.4853.

[54] J. N. Constantino and C. P. Gruber, *The Social Responsiveness Scale*, Los Angeles: Western Psychological Services, 2002, doi: https://doi.org/10.1007/978-1-4419-1698-3_296.

[55] M. Rutter, A. Bailey, and C. Lord, *The social communication questionnaire: Manual*, Los Angeles: Western Psychological Services, 2003.

[56] K. Gwet, "Inter-rater reliability: dependency on trait prevalence and marginal homogeneity," *Stat. Methods Inter-Rater Reliab. Assess.*, vol. 2, pp. 1-9, 2002.

[57] Y-W. Chang, C-J. Hsieh, K-W. Chang, M. Ringgaard, and C-J. Lin, "Training and testing low-degree polynomial data mappings via linear SVM," *J. Mach. Learn. Res.*, vol. 11, pp. 1471-1490, Aug. 2010, doi: 10.5555/1756006.1859899.

[58] L. Zhang, Z. Warren, A. Swanson, A. Weitlauf, and N. Sarkar, "Understanding performance and verbal-communication of children with ASD in a collaborative virtual environment," *J. Autism Dev. Disord.*, pp. 1-11, 2018, doi: 10.1007/s10803-018-3544-7.

[59] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, Burlington, Massachusetts, USA: Morgan Kaufmann, 17th November 2016.

[60] A. Tewari, T. Brown, and J. Canny, "A question-answering agent using speech driven non-linear machinima," in *Int. Conf. Intell. Virtual Agents*, Aug. 2013, pp. 129-138, doi: 10.1007/978-3-642-40415-3_11.

[61] R. Yaghoubzadeh, K. Pitsch, and S. Kopp, "Adaptive grounding and dialogue management for autonomous conversational assistants for elderly users," in *Int. Conf. Intell. Virtual Agents*, 2015, pp. 28-38, doi: 10.1007/978-3-319-21996-7_3.

CHAPTER 3: C-HG: A COLLABORATIVE HAPTIC-GRIPPER FINE MOTOR SKILL TRAINING SYSTEM FOR CHILDREN WITH AUTISM SPECTRUM DISORDERS

3.1 Abstract

Computer-assisted systems can provide efficient and engaging ASD intervention environments for children with Autism Spectrum Disorder (ASD). However, most existing computer-assisted systems target only one skill deficit (e.g., social conversation skills) and ignore the importance of other areas, such as motor skills, that could also impact social interaction. This focus on a single domain may hinder the generalizability of learned skills to real-world scenarios, because the targeted teaching strategies do not reflect that real-world tasks often involve more than one skill domain. The work presented in this chapter seeks to bridge this gap by developing a Collaborative Haptic-gripper virtual skill training system (C-Hg). This system includes individual and collaborative games that provide opportunities for simultaneously practicing both fine motor skills (hand movement and grip control skills) as well as social skills (communication and collaboration) and investigating how they relate to each other. We conducted a usability study with 10 children with ASD and 10 Typically Developing (TD) children (8–12 years), who used C-Hg to play a series of individual and collaborative games requiring differing levels of motor and communication skill. Results revealed that participant performance significantly improved in both individual and collaborative fine motor skill training tasks, including significant improvements in collaborative manipulations between partners. Participants with ASD were found to conduct more collaborative manipulations and initiate more conversations with their partners in the post collaborative tasks, suggesting more active collaboration and communication of participants with ASD in the collaborative tasks. Results support the potential of our C-Hg system for simultaneously improving fine motor and social skills, with implications for impacts of improved fine motor skills on social outcomes.

3.2 Introduction

Impairments in executive function skills (EF), which include domains such as working memory, flexible thinking, and inhibition of impulses, are commonly reported among individuals with Autism Spectrum Disorder (ASD) [Hill 2004; Ozonoff and Jensen 1999; Ozonoffetal. 1994]. Executive function skills are crucial for effective coordination and completion of tasks, with EF deficits associated with poor adaptive, academic, and employment outcomes [McLean et al. 2014; Biederman et al. 2006]. These skills are especially important to multitasking, or the ability to plan, coordinate, and complete multiple tasks within a given time period not in a sequential fashion but rather by interweaving tasks by switching back and forth between them [Law et al. 200]. Because of the relevance of EF to so many academic, adaptive, and employment tasks, several recent studies have investigated EF in ASD by using multitasking paradigms

[Mackinlay et al. 2006; Rajendran et al. 2011; White et al 2009]. Multitasking involves many aspects of EF that may be hard for someone with ASD, such as switching attention between components and practicing flexible thinking. Mackinlay et al. [2006] investigated multitasking in 14 children with high-functioning ASD (HF-ASD) and 16 typically developing (TD) controls. Results indicated that the ASD group generated significantly fewer strategic plans, attempted fewer tasks, and less flexibility in switching between tasks, and that they broke rules more frequently than the TD controls. Rajendran et al. [Rajendran et al. 2011] used a modified version of the Virtual Errands Task (VET) [McGeorge et al. 2001] to investigate EF and multitasking in 18 adolescents with HF-ASD and 18 TD controls. They found that inflexible planning, low inhibition, as well as difficulties with prospective memory (i.e., remembering to carry out intentions) may underlie multitasking difficulties in ASD. Finally, Hutchison et al. [2019] examined EF in relation to basic functional communication (FC) and more complex verbal conversation (VC) skills among 92 children with ASD and 94 TD controls. They reported that metacognition or "thinking about thinking" was a strong predictor of FC, while the domains of behavioral regulation and inhibition were predictive of VC skills. Therefore, they suggested that targeting EF domains specifically might improve FC and VC skills in children with ASD. Collectively, this work supports the importance of EF when understanding how to support people with ASD in learning new skills. The impact of EF on multitasking and its impact on individuals with ASD becomes more salient when one considers that multitasking is a ubiquitous requirement of everyday activities, including social interactions. Each interaction draws upon not only the need to inhibit impulses, pay attention to cues, remember what has just happened, and plan for what happens next; many also involve motor skills, another common area of deficits for many people on the autism spectrum [Rao et al. 2008]. Motor skill deficits (including gross motor, i.e., postural control and limb movements; and fine motor, i.e., object control, manual dexterity and visuo motor integration) are incredibly prevalent among children with ASD, and are estimated to occur in 90% of individuals with ASD across the lifespan [Gowen and Hamilton 2013; Ming et al. 2007;Jansiewicz et al. 2006; Forti et al. 2011; MacDonald et al. 2013]. Fine motor challenges in children with ASD include trouble with grasping and reaching [David et al. 2012], eye-hand coordination [Crippa et al. 2013], and handwriting skills [Johnson et al. 2013]. These basic tasks, which many people execute almost automatically, likely require extra effort, control, and mental attention from people with ASD. When thinking about designing intervention paradigms for individuals with multiple areas warranting attention, it follows that focusing on a single area of deficit (e.g., conversation) rather than incorporating multiple integrated targets (e.g., conversation as part of game-playing) may hinder the generalizability of learned skills to complex real-world activities. In particular, it would be problematic if additional real-word components created added levels of difficulty and complexity that could impact success. Existing literature suggests that teaching motor skills to children with ASD may help create a context for practicing social skills and lead to further social success [MacDonald et al. 2013]. For example, Chetcuti et al. [2019] conducted a study with 35 children with ASD and 20 TD children to examine the role of social motivation and motor execution factors in object-directed imitation difficulties in ASD. They found that difficulties in object-directed imitation in ASD might be the result of motor execution difficulties, and not reduced social motivation. Srinivasan et al. [2015] evaluated the impacts of rhythm, robotic and standard-of-care interventions on 36 children with ASD (5–12 years of age). They found that socially embedded movement-based contexts are valuable in promoting imitation/praxis, interpersonal synchrony and motor performance. Fulceri et al. [2018] applied Artificial Neural Networks (ANNs) to understand the entire spectrum of the relationship between motor skills and clinical variables. Their findings suggested that poor motor skills were a common clinical feature of preschoolers with ASD, relating both to the high level of repetitive behaviors and to the low level of expressive language. Collectively, these findings suggest that to benefit from social skills training, some individuals might also require training in other, related functional domains, like motor skills. Compared to extensive research focusing on the social deficits of ASD, motor deficits of children with ASD and their impacts on social skills are relatively underexplored, especially within the context of technological intervention. To address this issue, a training system that can provide combined social skill and fine motor skill practice is needed. One computer-assisted approach to address social challenges of children with ASD is the Collaborative Virtual Environment (CVE) [Zhao et al. 2018; Zheng et al. 2017; Battocchi et al. 2009]. Battocchi et al. [2009] designed a tabletop collaborative puzzle game featuring enforced collaboration to facilitate cooperative behaviors in children with ASD. However, their work did not consider the importance of haptic communication or motor skills for strengthening social interaction. Simulating the sense of touch and physical contact with shared objects in CVEs could enhance feelings of social presence and task performance [Sallnäs 2010; Basdogan et al. 2000]. Most existing work combining technology and fine motor skill intervention has focused not on social interaction, but on handwriting analysis [Rosenblum et al. 2016; Palsbo et al. 2012; Kim et al. 2013; Zhao et al. 2018; Zhao et al. 2018]. Although important to functional outcomes, handwriting does not capture the other fine motor tasks related to social interaction for children, such as holding and moving a puzzle piece, coloring a picture, or playing a video game. If children struggle with the fine motor components of these tasks, then it stands to reason that they may have less cognitive attention and energy left to focus on maintaining positive social interactions. There is, to our knowledge, no existing system that provides and measures responses to opportunities to simultaneously practice social skills (e.g., communication and collaboration skills) and fine motor skills (e.g., hand movement and grip control skills), or investigates the impact of fine motor skill improvement on social skill trajectory. The primary contributions of our work are two-fold. First, we developed a Collaborative Haptic-gripper virtual reality system (C-Hg) to simultaneously practice social skills (i.e., communication and collaboration skills) and fine motor skills (i.e., hand movement and grip control skills). In the collaborative mode, a Collaborative Haptic Virtual Environment (CHVE) provides flexible haptic interactions over the Internet between remote users. Social communication and cooperation skills are required to successfully implement the carefully designed collaborative fine motor tasks. Second, we conducted a usability study to explore the impact of our system on fine motor skills and, subsequently, the impact of fine motor skill performance on the change in social skill performance across tasks.

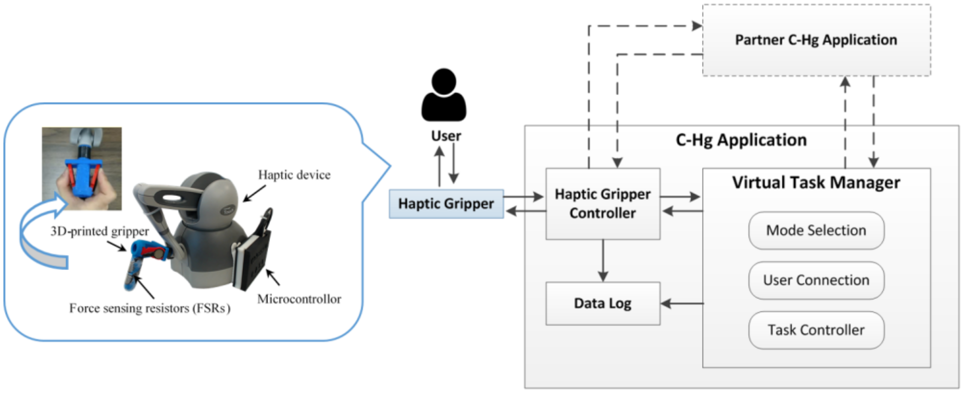


Figure 3-1: C-Hg System Architecture

3.3 C-Hg System Design

The Collaborative Haptic-Gripper virtual reality system (C-Hg) was developed from a prototype we used in a previous study to provide virtual fine motor tasks for single users [Zhao et al. 2018]. By incorporating a collaborative mode in which participants work together, C-Hg allows remote users to collaborate when performing virtual fine motor tasks. Fig. 3-1 shows the system architecture with the major modules. The user interacts with C-Hg via a customized interactive tool, called Haptic Gripper, with which the user can manipulate the virtual objects as well as feel the force feedback. The Haptic Gripper was constructed by augmenting a commercial haptic device, the Geomagic Touch Haptic Device1 with a 3D-printed gripper embedded with force sensing resistors (FSRs). The design details of the Haptic Gripper were previously presented by Zhao et al. [2018]. Two basic hand manipulations are provided by the Haptic Gripper. The first one is Movement Manipulation. For example, the user can change the position of the controlled virtual objects by holding and moving the gripper. The second one is Grip Manipulation, which requires the user to squeeze the red plates of the gripper using the fingers. This manipulation changes some property of the virtual object (e.g., the size of the object). The Haptic Gripper Controller obtains the grip force data and the hand position data of the user, and sends the data to the Virtual Task Manager to update the states of the virtual objects. The user then receives visual, auditory or force feedback in response to their movement and/or grip manipulation. The C-Hg system provides an individual training mode as well as a collaborative training mode.

In the individual training mode, the user performs the tasks alone to practice individual fine motor skills. In the collaborative training mode, two users complete tasks in a cooperative manner that can foster their communication and collaboration skills within the context of fine motor tasks. To implement the collaborative training mode, we tested the performance of three different control architectures: a centralized control architecture, a distributed control architecture, and a wave-variable-based control architecture [Zhao 2021]. We found that the centralized control architecture and the wave-variable-based control architecture had better performance with regard to frequency response, position error and force rendering. As the time delay increased, the wave-variable-based control architecture outperformed the centralized control architecture. For the usability study, we invited pairs of participants to a study session where each participant sat in a different room of the same building. Since the study was conducted in a LAN environment with small network delays, we used the centralized control architecture, which required lower-complexity implementation compared to the wave-variable-based control architecture. When two users selected the collaborative mode and wanted to play together, their applications were connected. One of these applications would play the role of a server to manage all input data to update the task states and to synchronize the updated task states on both applications. The virtual tasks of the C-Hg system were carefully designed (as discussed later) to assess and train the user's fine motor skills that require the coordination of the finger, hand, arm, and eye. The Data Log records specific user data and task data, which are time-synchronized for offline processing and analysis.

3.3.1 Fine Motor Virtual Tasks

We designed three types of fine motor virtual tasks separately focusing on Grip Manipulation ("*Curling*" task), Movement Manipulation ("*Go, Wheel!*" task), and complex manipulation that involves both Grip manipulation and Movement manipulation ("*Prize Claw*" task and "*Green Path*" task). These tasks were designed in consultation with ASD clinicians with the objective of providing both fine motor manipulation and collaboration opportunities through embedded rules that must be followed for success. In the collaborative games, two users were required to control one object together, but each had a different responsibility. Instead of taking turns to control the object, users were expected to pay more attention to the partner and had more communications and interactions with each other. All the tasks had time limits, and were required to finish within a certain period of time (e.g., 3 minutes). The task performance was evaluated by the task score.

Grip Task: *Curling* Many of children's activities require grip manipulation skill, such as handwriting and typing. Studies have shown that children with ASD might exhibit inappropriate grip force adjustment, and greater grip force variability compared to their TD peers [Wang et al. 2015]. The "*Curling*" task was designed to practice grip control capability.

As shown in Fig. 3-2, the *Curling* task requires the user to move a curling stone with the goal to stop it within a circular target with rewards. The user applies controlled pressure on the gripper in order to move the curling stone along the *y* axis, and releases the gripper to make it to decelerate to a stop. The motion equations of the curling stone are:

$$a = \frac{F - f}{m}$$

$$y_{curr} = y_{last} + v_{last} \times \Delta t + \frac{1}{2} \times a \times \Delta t^2 \qquad (2)$$

$$v_{curr} = v_{ast} + a \times \Delta t$$

where *F* is the grip force the user applies on the gripper, *f* is the dynamic friction, *m* is the mass of the curling stone, y_{last} and v_{last} are the last position and velocity of the curling stone, respectively, and y_{curr} and v_{curr} are the new current position and velocity of the curling stone, respectively. The position and speed of the curling stone reflects pressure applied by the user. An arrow is also provided to indicate the move direction of the curling and the user's applied pressure. The user should estimate the stopping distance of the curling stone according to its velocity and position, and carefully adjust the grip force to make the curling stone remain within the target area. When the user successfully places the curling stone within the target, or when the user moves the curling stone over the target, the target disappears and a new one appears at a random location along the *y* axis. To enhance user immersion, a constant resistance force feedback is generated through the haptic device to simulate the feeling that the user may get when pushing the curling stone.

In the collaborative mode, two users move the curling stone together, but along two perpendicular axes. As shown in Fig. 3-2, Player A can only move it along the *x* axis, while Player B can only move it along the *y* axis. Their joint force determines the movement direction and speed of the curling stone, and they must collaborate if they want to move along the diagonal. One user cannot move the curling stone to the target without the help of the partner. For example, in the collaborative task of Fig. 3-2, if Player A and Player B apply the same pressure on the gripper at the same time, the curling stone can move straight towards the target. But if Player A applies bigger pressure or only Player A applies pressure on the gripper, the curling stone will move toward the left side of the target (close to the *x* axis). The position and speed of the curling stone reflects the pressures applied by the players, based on which two players can discuss when to grip, who should grip, and how much pressure to use when controlling the curling stone together. Therefore, the most effective grip- and movement-based manipulations for getting more rewards have to be negotiated and performed in a cooperative way.

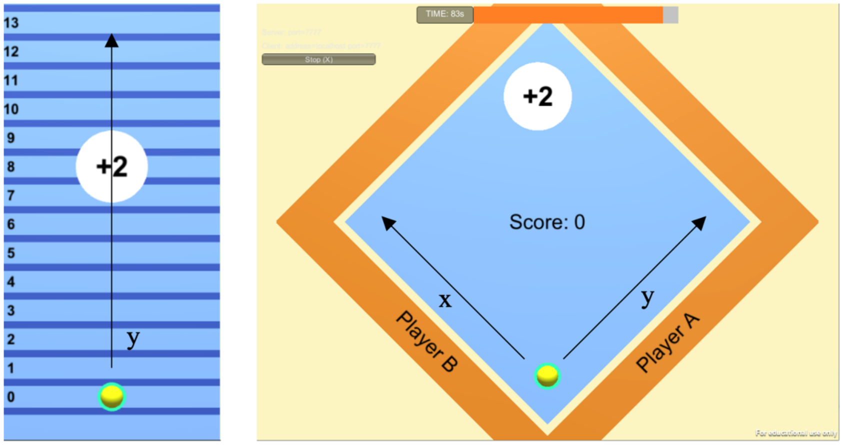


Figure 3-2: The grip task *Curling* in the individual mode (left) and in the collaborative mode (right).

Movement Task: *Go, wheel!* Children with ASD show differences in fine motor skills from their TD peers regarding movement speed, movement adjustment, and eye-hand coordination [Anzulewicz et al. 2016, Cook et al. 2013]. The "*Go, wheel!*" task was designed for the user to practice movement control.

In the individual "*Go, Wheel!*" task, the user is required to pull a rolling wheel to the left side or to the right side to collect gold coins while avoiding rocks on the road. Once the wheel is pulled to one side or hits a rock, it falls to the ground at a certain speed unless the user successfully pulls it back. The tilting angle of the wheel is controlled by the following equations:

$$\alpha_{curr} = \begin{cases} \arcsin\dfrac{d \times \sin\alpha_{last} + disp_x}{d}, & if \quad disp_x > 0 \\ \alpha_{last} + \alpha_0 \times \Delta t, & if \quad \alpha_{last} \neq 0 \; and \;\; disp_x = 0 \end{cases} \tag{3}$$

where $disp_x$ is the displacement of the user's hand movement along the *x* axis, *d* is the diameter of the wheel, α_0 is the falling angle per unit of time, α_{last} is the last tilting angle of the wheel, and α_{curr} is the new tilting angle of the wheel. When the wheel is falling to one side or the user is pulling the wheel, a spring force feedback is generated to simulate the pulling force, which prompts the user to take actions.

In the collaborative mode, two users control the wheel together, but each user can only pull the wheel to one side. As shown in Fig. 3-3, Player A can only pull it to the right side, while the Player B can only pull it to the left side. When they want to get a gold coin on the right side, only Player A can pull the wheel to the right side to get it, and only Player B can pull the wheel back to prevent it from falling on the ground. Thus, each manipulation for catching a coin requires that the two players collaborate, in that each must pay attention to the other's action and adjust at the right time in order to keep the wheel upright. The task also requires communication to decide which coin to catch, and when and who should pull the wheel, since the coins can come from either left side or right side, and at different speeds. If two players pull the wheel to the opposite sides, they would fail to get the coin. Therefore, the two players have to communicate in order to coordinate their movements, catch more coins, and get a higher score.

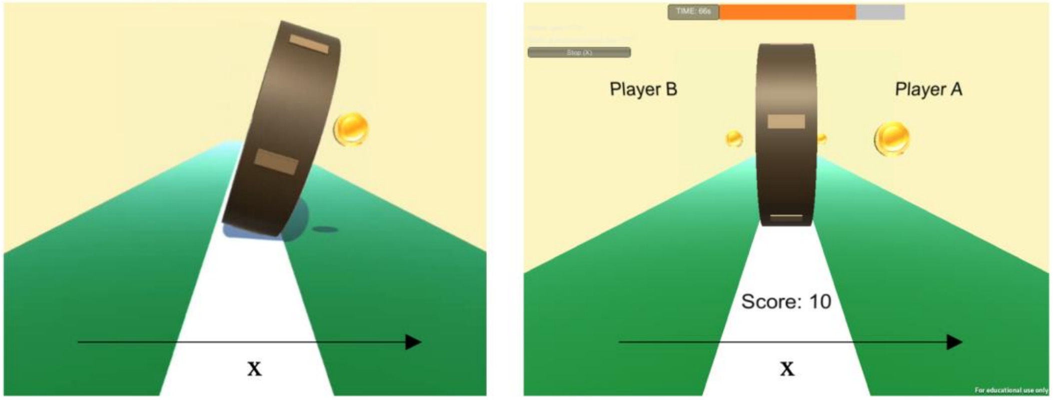


Figure 3-3: The movement task *Go, wheel!* in the individual mode (left) and in the collaborative mode (right).

Complex Task: *Prize Claw* In the individual "*Prize Claw*" task (Fig. 3-4), the user is required to perform both grip and movement manipulations. The task is similar to the *Prize Claw Machine* in the real world. The user moves the claw at the top of the machine. In order to catch a prize successfully, the user should move the claw to the position right above the prize, then grip to make the claw go down in an attempt to grasp the prize. The shadow position of the claw would indicate whether the claw is just above the prize or not, based on which the user can decide the move direction. The user must then move the claw to put the prize into the hole to get rewards. The states of the claw are controlled by the following equations:

$$
\begin{gathered}
x_{curr} = x_0 + disp_x, \qquad if\ y_{last} = y_{top} \\
z_{curr} = z_0 + disp_z, \qquad if\ y_{last} = y_{top} \\
des_y = \begin{cases} y_{top}, & if\ C = true \\ y_{middle}, & if\ C = false\ and\ F \in MediumRange \\ y_{bottom}, & if\ C = false\ and\ F\ \in LargeRange \end{cases} \\
y_{curr} = y_{last} + v_y \times \Delta t, \qquad if\ y_{curr}\ \neq des_y
\end{gathered} \tag{4}
$$

where *disp*$_x$ and *disp*$_z$ are the displacements of the user's hand movement along the x and z axis, respectively; x_0, z_0 and y_{top} are the initial positions of the claw, and x_{curr}, y_{curr} and z_{curr} are the new positions of the claw. The *des*$_y$ is the destination position of the claw along the y axis determined by the applied grip force (F) and the state whether a prize is being caught (C), and v_y is the speed at which the claw moves to the destination position. Greater force makes the claw go deeper to catch the prizes at the bottom, while a smaller force can make the claw catch the prizes at the middle height. When the claw grasps a prize, the claw will take the prize back to the top, and then the user can move it to the hole's location. The user should keep squeezing the red plates of the gripper (see Fig. 3-1) to maintain the required force in order to hold the prize before dropping it. When the user stops gripping, the prize falls into the hole or on the table.

The user is restricted to moving the claw within the prize claw machine. To improve immersion, a viscous effect and a friction effect accompany the movement of the claw. A resistance force feedback is generated when the claw collides with the glass enclosure of the machine. The user can also feel the change in weight when he/she picks up the prize or drops it into the hole.

In the collaborative mode, game play is portrayed as if two users are standing at adjacent sides of the prize claw machine. Users control the claw together. They have different game views (e.g., the giraffe is on the right side from the player B's viewpoint, while it is on the left side from the player A's viewpoint in Fig. 3-4). Users can only move the claw side along the plane that corresponds with their viewpoints. To pick up a prize, users must communicate and share information from their respective viewpoints about the claw position relative to the prize position. Then, they must both grip at the same time to pick up the prize, and keep gripping to hold the prize until it is over the hole, at which point they release the red plates of the grippers to drop the prize into the hole. If only one player grips, the claw will not go down to catch the prize. If either one of them releases the gripper before reaching the hole, the caught prize will drop on the table, and they would have to pick up it again. Thus, the users must communicate with one another to determine which prize to pick and whether they are ready to pick it up and drop it off in order to get as many prizes as possible within limited time.

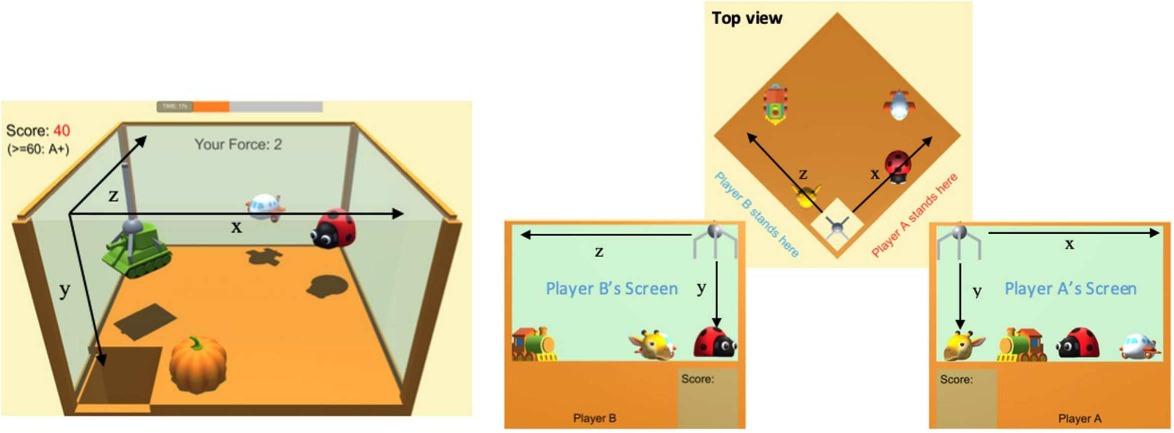


Figure 3-4: The complex task *Prize Claw* in the individual mode (left) and in the collaborative mode (right).

Complex Task: *Green Path* Similar to the "*Prize Claw*" task, the "*Green Path*" task requires both grip and movement manipulations. As shown in Fig. 3-5, the user picks up a yellow ball from the upper left corner of the board, and then moves it to the bottom right corner through a "safe" path (denoted by the green color). The user can manipulate the size of the ball by adjusting their grip strength. They must change and match the size of the ball with rings placed along the path in order to win rewards. The states of the ball are controlled by the following equations:

$$
\begin{gathered}
x_{curr} = x_0 + disp_x, \qquad if\ F > SmallRange \\
z_{curr} = z_0 + disp_z, \qquad if\ F > SmallRange \\
y_{curr} = \begin{cases} y_{top}, & if\ F > SmallRange\ and\ G = true \\ y_{bottom}, & if\ F \in SmallRange\ and\ G = true \\ y_{last} - at^2, & if\ G = false \end{cases} \\
S_{curr} = \begin{cases} S_{small}, & if\ F \in LargeRange \\ S_{big}, & if\ F \in MediumRange \end{cases}
\end{gathered} \tag{5}
$$

where y_{top} and y_{bottom} are the *y* position of the ball when it is picked up and when it is dropped down on the green path, respectively; *a* is the falling acceleration when the ball moves outside the green path, *G* represents if the ball is within the green path, S_{small} and S_{big} are the ball size when the user grips hard and when the user grips lightly, respectively. Like the *Prize Claw* task, the user can feel the change in weight when the ball is picked up and when it is dropped off.

In the collaborative mode, two users work together to move the yellow ball and obtain as many rewards as they can. Each user controls a movement handle. The horizontal handle only moves along the *y* axis, while the vertical handle only moves along the *x* axis. The crossing point (represented by the red dot in Fig. 3-5) of the two handles is the “control dot” that can pick up, move and drop the yellow ball. The ball can only be picked up and moved when the control dot is above it and both users are gripping. Since one user only controls one movement direction of the ball, two users have to work together to move the ball to the destination. To promote the need for communication between users in the collaborative mode, the green path and the reward rings are visible to one user but invisible to the other user. When one user can see the green path, the other one can see the reward rings. Therefore, to be successful, users must share the position information of the green path and the reward rings, and together decide where and when to move the ball or adjust the size of the ball, in order to avoid dropping the ball into the blue areas or missing the rewards on the path.

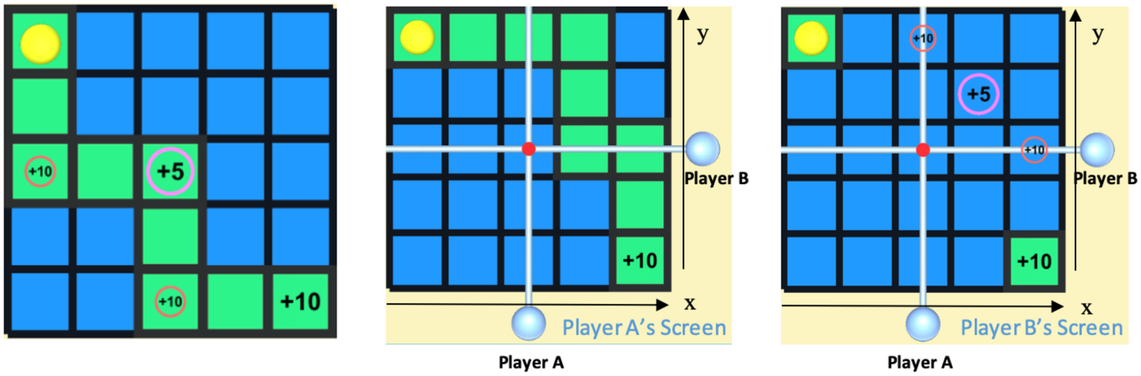


Figure 3-5: The complex task Green Path in the individual mode (left) and in the collaborative mode (right).

3.4 Usability Study

We conducted a usability study to   To achieve these goals, we recruited 10 pairs of ASD and TD participants. In pre- and post tests, each participant completed the individual fine motor tasks to assess baseline and post-session fine motor skill performance, and worked with his/her partner to complete the collaborative fine motor tasks to evaluate the impact of fine motor skill training on their social skills. This study was approved by Vanderbilt University Institutional Review Board (IRB).

3.4.1 Participants

We recruited 10 children with ASD and 10 TD children (ages: 8-12 years, mean age: 10.85 years) through an existing clinical research registry. We then created 10 age- and sex-matched pairs (ASD-TD), who perform the fine motor tasks together in the collaborative mode in order to evaluate the impact of individual fine motor skill training on their social skills. Table 3-1 shows the participant characteristics. As seen in Table 3-1, autism symptoms were indexed using the Autism Diagnostic Observation Schedule-Second Edition (ADOS-2) [Lord 2000], the Social Responsiveness Scale, Second Edition (SRS-2) [Constantino and Gruber 2007], and the Social Communication Questionnaire – Lifetime score [Rutter et al. 2003].

Table 3-1: Participant Characteristics

Metrics	TD Group (N = 10)	ASD Group (N = 10)
	Mean (SD)	Mean (SD)
Age (Years)	10.85 (1.86)	10.94 (1.36)
SRS-2 total raw score	23.5 (25.61)	103.25 (21.76)
SRS-2 T score	48.6 (15.05)	79 (8.14)
SCQ Lifetime total score	2.5 (2.59)	21.75 (7.38)
ADOS-2 total score	/	17.88 (3.97)

3.4.2 Procedure

Each participant pair came for three times separated by approximately one week (Fig. 3-6). Each participant pair was required to complete a pre-test and an individual training session in their first visit, two individual training sessions in their second visit, and an individual training session and a post-test in their third visit. As described in Section 3.3, all the tasks have the individual mode and collaborative mode. Participants performed the individual tasks independently in the individual mode, while they needed to connect with their paired partner's computer to perform the collaborative tasks in the same virtual task environment in the collaborative mode.

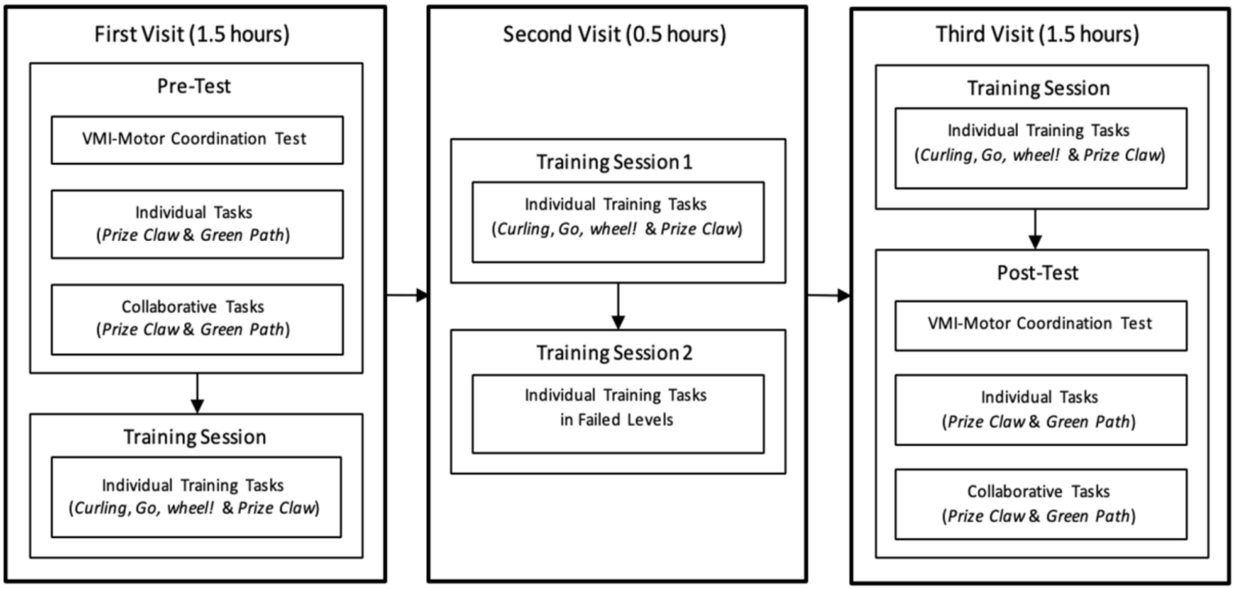

First Visit (1.5 hours)
Pre-Test
VMI-Motor Coordination Test
Individual Tasks (*Prize Claw* & *Green Path*)
Collaborative Tasks (*Prize Claw* & *Green Path*)
Training Session
Individual Training Tasks (*Curling, Go, wheel!* & *Prize Claw*)
Second Visit (0.5 hours)
Training Session 1
Individual Training Tasks (*Curling, Go, wheel!* & *Prize Claw*)
Training Session 2
Individual Training Tasks in Failed Levels
Third Visit (1.5 hours)
Training Session
Individual Training Tasks (*Curling, Go, wheel!* & *Prize Claw*)
Post-Test
VMI-Motor Coordination Test
Individual Tasks (*Prize Claw* & *Green Path*)
Collaborative Tasks (*Prize Claw* & *Green Path*)

Figure 3-6: The Study Procedure.

In visits 1 and 3, the pre- and post-tests consisted of a mix of virtual and real-world tasks. To measure fine motor performance in real-world tasks, we used the Motor Coordination subtest of the Beery-Buktenica Developmental Test of Visual-Motor Integration (VMI Motor Coordination Test) [Beery and Beery 2010]. Regarding virtual tasks, the four individual tasks (two *Prize Claw* and two *Green Path*) and four collaborative tasks (two *Prize Claw* and two *Green Path*) were designed to assess the within system fine motor skills (i.e., hand movement and grip control skills) that were practiced in the individual training sessions as well as the social skills (i.e., communication and collaboration skills) that were required in the collaborative tasks.

In the individual training sessions, each participant completed a grip task (*Curling*), a movement task (*Go, wheel!*) and a complex task (*Prize Claw*) independently. These tasks were developed with three levels of difficulty, and the system adaptively adjusted the task difficulty level according to the participant's real-time performance (Table 3-2). The participants started from the easiest level (Level 1). If they achieved the goal within the given task time (1 minute), they entered the next, harder level. Otherwise, they either had to go back to the lower level or stay at the lowest level (Level 1) to have more practice. Each task ended either when the participant achieved the goal in Level 3 or after the participant finished five trials.

For the *Curling* task, the task level was developed by decreasing the size of the circular target to increase the task difficulty. For the *Go, wheel!* task, the golden coins move with the same speed in Level 1 and with varying speeds in Level 2. In Level 3, some rocks appear on the road that the participant needs to avoid in order to maintain higher scores. For the *Prize Claw* task, all the prizes are randomly placed on the table and the participant needs to apply greater force to grab them in Level 1. In Level 2, some prizes stay in the air and the participant needs to apply light grip force to catch them. In Level 3, all the prizes are moving on the table and the participant needs to carefully consider where and when to put down the claw in order to catch the moving prizes.

As shown in Fig. 3-6, in the first visit, each participant completed the pre-test and one training session. In the second visit, they completed Training Session 1 and Training Session 2. Training Session 1 was identical to the training sessions provided in the first visit and second visit, which provided the training tasks in the adaptive way as described in the previous paragraph. As each participant completed Training Session 1, the system recorded the level at which each participant failed the task to determine the participant's maximum task difficulty level. Then, in Training Session 2, each participant only took the tasks in the failed level. For example, if one participant successfully completed Level 1 of *Curling* Task, but failed to complete Level 2, Level 2 was the failed level of *Curling* Task for that participant. Thus, in Training Session 2, the participant only took the *Curling* Task in Level 2 that was challenging for the participant. If one participant succeeded at all levels in Training Session 1, he/she would not need to take

Training Session 2. Finally, in the third visit, participants completed one training session and then the post-test.

Table 3-2: Training Task Configurations.

Tasks	Level 1	Level 2	Level 3
Curling	Large target	Medium target	Small Target
Go, wheel!	Constant target speed	Varying target speed	Varying target speed, Rock obstacles
Prize Claw	All targets stay on the table	Some targets stay in the air	All targets move on the table

3.5 Results and Discussions

Participants provided feedback about their experiences using the system by completing a paper-and-pencil questionnaire at the end of their third visit. Due to the small sample size, we used the Wilcoxon signed-rank test to compare participants' pre- and post-test performance [Wilcoxon 1945]. We also reported *r*, the effect sizes with the significance cutoffs of large (>0.5), medium(>0.3) and small (>0.1) [Cohen 1988].

3.5.1 Acceptability of the C-Hg System

We collected participants' feedback about the Haptic Gripper and virtual games with a 5 point-Likert scale questionnaire including nine questions. Table 3-3 shows the survey questions as well as the participants' responses. The responses to questions 1-5 indicated that the interactive Haptic Gripper device was well accepted by the participants. It was not difficult for the participants to learn how to use the Haptic Gripper and to perform manipulations through it in the game. Most of them expressed interest in using this device with haptic effects to play video games. They also reported that the haptic effects were useful to help them understand their manipulations. The responses to questions 6-7 indicated that the virtual games were easy to understand and the participants enjoyed these games. The responses to questions 8-9 indicated that the virtual games could promote communication and collaboration between partners. All these results suggested that C-Hg system was well designed to maintain the participant's interests on the virtual tasks as well as to provide social interaction opportunities. The *t*-tests did not reveal significant differences in mean responses across groups (all *p* values > .05).

Table 3-3: Survey Feedback of TD Group (N = 10) and ASD group (N = 10)

No.	Questions	TD	ASD	*t*	*p*
		Mean(SD)	Mean(SD)	*(18)*	
1	How difficult was it to use the Haptic Gripper? (5-very difficult, 1-very easy)	2.5 (0.97)	2.5 (0.97)	0.0	1.0
2	How much did you like to use the Haptic Gripper to play games? (5-very much, 1-not at all)	3.8 (1.03)	3.3 (1.42)	0.90	.379
3	How much did you like these haptic effects? (5-very much, 1-not at all)	3.9 (0.88)	3 (1.15)	1.96	.065
4	How useful are these haptic effects to help you understand your operations? (5-very useful, 1-absolutely useless)	3.4 (0.97)	3.3 (1.49)	0.12	.909
5	How did you do to manipulate the Haptic Gripper? (5-excellent, 1-very bad)	3.8 (0.95)	3.4 (1.07)	0.95	.355
6	How much did you like these games? (5-very much, 1-not at all)	4 (1.25)	3.8 (1.14)	0.38	.712
7	How difficult was it to understand how to play these games? (5-very difficult, 1-very easy)	1.9 (0.99)	2.2 (1.14)	-0.6	.538
8	How important was it to talk to your partner in order to win the collaborative games? (5-very important, 1- absolutely useless)	4.4 (0.70)	4.1 (1.20)	0.68	.503
9	How important was it to work together with your partner in order to win the collaborative games? (5-very important, 1- absolutely useless)	4.3 (1.25)	4.3 (0.95)	0.0	1.0

3.5.2 Performance Improvement

Table 3-4 presents the performance results of both the TD group and the ASD group in the individual fine motor skill training tasks. Both groups achieved statistically significant improvements on the VMI Motor Coordination test (*VMI*) with large/medium effect sizes (TD: *Relative Change* (*RC*) = 6.57%, *Z* = -2.26, *p* = 0.024, *r* = *-0*.51; ASD: *RC* = 15.38%, *Z* = -2.21, *p* = 0.027, *r* = *-0*.49). According to the VMI standard score interpretation [Beery and Beery 2010], the performance of the TD group remained at the average level (90-109) in both pre-test (99) and post-test (105.5), while the ASD group was at the below-average level (80-89) in the pre-test (84.5) and reached the average level in the post-test (97.5). As for the individual virtual fine motor tasks, both groups showed statistically significant improvements with large effect sizes

(TD in **Ind**ividual ***P**rize **C**law* tasks: *RC* = 22.2%, *Z* = -2.67, *p* = 0.008, *r = -0*.60; ASD in **Ind**ividual ***P**rize **C**law* tasks: *RC* = 61.5%, *Z* = -2.69, *p* = 0.007, *r = -0*.60; TD in **Ind**ividual ***G**reen **P**ath* tasks: *RC* = 44.3%, *Z* = -2.80, *p* = 0.005, *r = -0*.63; ASD in **Ind**ividual ***G**reen **P**ath* tasks: *RC* = 26.1%, *Z* = -2.71, *p* = 0.007, *r = -0*.61). It is to be noted that the "*Green Path*" tasks were not provided in the training sessions, and only provided in the pre- and post-tests. However, participants still showed significant performance improvements in the "*Green Path*" tasks. All these results suggested that the individual fine motor skill training tasks had a positive impact on participant fine motor skills.

Table 3-4: Individual Performance Results

	TD Group (N = 10)						ASD Group (N = 10)					
Metrics	Pre	Post	*RC*[d]	*Z*	*p*	$\lvert r \rvert$	Pre	Post	*RC*	*Z*	*p*	*r*
	Mdn	Mdn	*(%)*				Mdn	Mdn	*(%)*			
VMI[a]	99	105.5	6.57	-2.26	.024*	-.51†	84.5	97.5	15.4	-2.21	.027*	-.49
Ind PC[b]	4.5	5.5	22.2	-2.67	.008*	-.60†	3.25	5.25	61.5	-2.69	.007*	-.60†
Ind GP[c]	22	31.75	44.3	-2.80	.005*	-.63†	22	27.75	26.1	-2.71	.007*	-.61†

[a], scores on the VMI Motor Coordination test
[b], individual *Prize Claw* task score
[c], individual *Green Path* task score
[d], *relative change* computed by (post - pre)/pre *100%
**p<.05*, †|*r*|>*0.5*

3.5.3 Impact of Fine Motor Skill Training on Social Skills

The collaborative performance results are shown in Table 3-5. Both groups showed statistically significant improvement in performance on the collaborative tasks (***Col**laborative **P**rize **C**law* tasks: *RC* = 9.5%, *Z* = -2.55, *p* = 0.011, *r = -0*.57; ***Col**laborative **G**reen **P**ath* tasks: *RC* = 16.75%, *Z* = -2.8, *p* = 0.005, *r = -0*.63). The Spearman correlation analysis also found strong/medium correlations between the performance results of individual tasks and collaborative tasks (ASD in ***P**rize **C**law* tasks: *ρ* = 0.498, *p* = 0.025; TD in ***P**rize **C**law* tasks: *ρ* = 0.208, *p* = 0.379; ASD in ***G**reen **P**ath* tasks: *ρ* = 0.427, *p* = 0.061; TD in ***G**reen **P**ath* tasks: *ρ* = 0.574, *p* = 0.008), which indicated that the improvements in collaborative tasks were related to improvements in individual tasks. These results suggest that individual fine motor skill practice led to better collaborative fine motor performance when required to use communication and collaborative skills for success.

Table 3-5: Collaborative Performance Results of TD-ASD Pairs (N = 20, Pairs = 10)

Metrics	**Pre**	**Post**	***RC*[g]**	***Z***	***p***	***r***
	Mdn	Mdn	*(%)*			
Col PC[a]	9.5	12.75	34.2	-2.55	.011*	-.57†

Col GP[b]	16.75	31	85.0	-2.80	.005*	-.63†
ASD PC Col Ratio[c]	0.49	0.63	26.6	-2.40	.017*	-.54†
TD PC Col Ratio	0.57	0.65	13.2	-1.58	.114	-.35
ASD GP Col Ratio	0.75	0.81	7.85	-1.99	.047*	-.44
TD GP Col Ratio	0.77	0.81	4.94	-1.27	.203	-.29
ASD Col PC INIT[d]	12	15.25	27.1	-0.77	.44	-.17
TD Col PC INIT	20.75	15.25	-26.5	1.27	.20	.28
ASD Col PC RESP[e]	10	6	-.4	0.49	.62	.11
TD Col PC RESP	8.47	7	-.2	0.35	.72	.08
PC Col INIT Switch[f]	15.25	16.25	6.56	0.49	.62	.11
ASD Col GP INIT	13.25	18.75	41.5	-2.14	.033*	-.48
TD Col GP INIT	17.75	18.5	4.23	-1.48	.139	-.33
ASD Col GP RESP	5.75	5.5	-4.35	0.41	.682	.09
TD Col GP RESP	5.75	4.75	-17.4	1.07	.282	.24
GP Col INIT Switch	14.5	17	17.2	-1.89	.059	-.42

[a], collaborative *Prize Claw* task score
[b], collaborative *Green Path* task score
[c], the ratio of collaborative operations to total operations
[d], the frequency of initiating a conversation
[e], the response frequency
[f], the switch frequency of conversation initiator
[g], *relative change* computed by (post - pre)/pre *100%
*p<.05, †|r|>0.5

We also found that all participants, whether with ASD or TD, more frequently performed collaborative manipulations in the collaborative tasks of the post-test. Specifically, participants with ASD significantly increased their collaborative manipulations in the collaborative tasks of post-test (***Col***_laborative_ ***Ratio*** of *ASD* in *collaborative* ***P***_rize_ ***C***_law_ tasks: *RC* = 26.6%, *Z* = -2.4, *p* = 0.017, *r* = -0.54; ***Col***_laborative_ ***Ratio*** of *ASD* in *collaborative* ***G***_reen_ ***P***_ath_ tasks: *RC* = 7.85%, *Z* = -1.99, *p* = 0.047, *r* = -0.44). These results indicated increased interaction and collaboration between partners in the collaborative tasks of post-test, and might suggest that improved fine motor skills would promote collaborative activities of participants.

To analyze the conversation pattern of participants, two human coders (trained graduate students with experience collecting and analyzing qualitative and quantitative data) were recruited to manually transcribe and code the conversation audio from 80 recorded videos of collaborative tasks (four hours of video data in total). We provided a framework that described the concrete definitions of types of utterances, and a few examples about how to code the conversation data for the human coders. Each human coder independently coded the same data. After each rater completed coding, an interrater agreement of 92.5% was found. For the coding that were not in agreement, two human coders reconciled differences via a consensus in which any discrepancies were discussed and resolved. The final set of results are based on the input from both raters.

As described in Section 2.2, the conversations between two players in all the collaborative tasks involved strategy discussion and information sharing. Timely communication was important for players to obtain higher scores. In order to evaluate how often one player communicated with his/her partner, we defined three types of utterances, *Initiation* (*INIT*), *Response* (*RESP*), and *Initiation Switch* (*INIT Switch*). *Initiation* represented one player's statement that started a conversation. *Response* represented one player's feedback to the partner's statement. *Initiation Switch* represented the switch of the conversation initiator from one player to the other one. Based on the final coding data, we calculated the frequency of *INIT*, *RESP*, and *INIT Switch* in each collaborative task. Table 3-6 shows a sample of three conversations recorded in a collaborative *Green Path* task. First, Player A started a conversation asking where to go, and Player B responded with the direction information (recall that path position was only visible to one player). Second, Player B started a new conversation to direct the movement. Then, Player A started another conversation to provide the reward information (recall that reward position was only visible to one player), and Player B responded with an acknowledgement. Therefore, there were two *INITs* of Player A, one *INITs* of Player B, and two *RESPs* of Player B. We also see that the conversation initiator switched from Player A to Player B, and then switched back to Player A. Thus, we counted the frequency of *INIT Switch* as 2.

Table 3-6: A sample of conversations recorded in a collaborative *Green Path* task.

No.	Player A	Player B
1	Now where will we go? <*Initiation*>	Left. <*Response*>
2		Right. Down. Right, right. <*Initiation*>
3	Wait, right here. There is a reward. <*Initiation*>	Okay. <*Response*>

As shown in Table 3-7, results indicated that in all the collaborative tasks of the pre-test and post- test, both the ASD and TD participants initiated more conversations than providing responses, which is reasonable since participants could respond to their partners with an action instead of an utterance. However, we found that the TD participants maintained a significant difference between initiations and responses, in both the pre-test and the post-rest (**TD** in ***Col****laborative* ***P****rize* ***C****law* tasks of ***Pre***-test: Z = 2.81, p = 0.005, r = -*0*.63; **TD** in ***Col****laborative* ***P****rize* ***C****law* tasks of ***Post***-test: Z = 2.66, p = 0.008, r = -0.59; **TD** in ***Col****laborative* ***G****reen* ***P****ath* tasks of ***Pre***-test: Z = 2.70, p = 0.007, r = -0.60; **TD** in ***Col****laborative* ***G****reen* ***P****ath* tasks of ***Post***-test: Z = 2.80, p = 0.005, r = -0.63), while the difference among ASD participants is only significant in the post-test (**ASD** in ***Col****laborative* ***G****reen* ***P****ath* tasks of ***Post***-test: Z = 2.30, p = 0.021, *r* = - 0.51; **ASD** in ***Col**laborative **G**reen **P**ath* tasks of ***Post***-test: *Z* = 2.80, *p* = 0.005, *r* = -0.63). The results showed in Table 3-5 also indicated that ASD participants made more initiations in the post-test, with significant improvements found in the ***Col**laborative **G**reen **P**ath* tasks (*Z* = -2.14, *p* = 0.033, *r* = -0.48). These results might suggest increased active communication among the ASD participants, and they tended to provide information or commands in the post-test.

These findings are consistent with other study results [Zhang et al. 2008, Owen et al. 2008]. Zhang et al. found that children with ASD increased question asking over the course of the peer-mediated training intervention. In addition, in the post-test, the switch frequency of who initiated a conversation increased (***INIT Switch*** in *collaborative **P**rize **C**law* tasks: *RC* = 6.56%, *Z* = 0.49, *p* = 0.62, *r* = -0.11; ***INIT Switch*** in *collaborative **G**reen **P**ath* tasks: *RC* = 17.2%, *Z* = -1.89, *p* = 0.059, *r* = - 0.42), though no significant difference was found, which also suggested that participants more actively and effectively communicated with their partners to share information or provide commands in the post-test collaborative tasks. Considering the improvements in task performance, these results might suggest that having the opportunity to practice fine motor skills could foster more active communication.

Table 3-7: The comparison between initiation frequency and response frequency among the ASD group (N = 10) and TD group (N = 10)

Metrics	*INIC*[c]	*RESP*[d]	*Z*	*p*	*r*
	Mdn	Mdn			
ASD Col PC[a] *PRE*	12	10	1.74	.082	.39
TD Col PC PRE	20.75	8.47	2.81	.005*	.63†
ASD Col PC POST	15.25	6	2.30	.021*	.51†
TD Col PC POST	15.25	7	2.66	.008*	.59†
ASD Col GP[b] *PRE*	13.25	5.75	1.38	.169	.31
TD Col GP PRE	17.75	5.75	2.70	.007*	.60†
ASD Col GP POST	18.75	5.5	2.80	.005*	.63†
TD Col GP POST	18.5	4.75	2.80	.005*	.63†

[a], collaborative *Prize Claw* task score
[b], collaborative *Green Path* task score
[c], the frequency of initiating a conversation
[d], the response frequency
**p<.05*, †|*r*|*>0.5*

3.6 Conclusion and Future Work

Many individuals with ASD have not only social-communication challenges, but also experience delays in their fine motor skills. Because fine motor performance is inherent to many social-communication tasks of childhood, such as playing a board game or completing a group assignment, these fine motor deficits may impact their ability to engage with real-world social activities. To provide opportunities for fine motor skill practice within an automated measurement system and collaborative peer context, we developed a collaborative haptic-gripper skill training system (C-Hg). By using a customized Haptic Gripper, the system can provide grip control training and hand movement training in both individual as well as collaborative modes. In the individual mode, the system can adaptively adjust the difficulty level of training tasks according to the user's real-time performance. In the collaborative mode, the user is required to communicate and collaborate with their partner to complete the training tasks together. These tasks simultaneously engage fine motor as well as social-collaboration skills in children with ASD while allowing us to investigate how they relate to each other.

This usability study with 10 children with ASD and 10 TD children indicated that this system was well accepted by all participants, regardless of diagnostic status. Both ASD children and TD children showed significant improvements on the VMI pre- and post-test, reflecting the relevance of the task to standardized real-world motor assessment tools. Participants in both groups also showed significant improvements in task performance during individual and collaborative tasks, which suggests that this system not only impacts individual fine motor skills, but also fosters social communication and collaboration. We also found strong/medium correlations between individual task performance and collaborative task performance, indicating that improved individual fine motor skills relate to better performance in social collaborative tasks. In addition, participants with ASD were found to conduct more collaborative manipulations and initiate more conversations with their partners in the post-test, suggesting that more active collaboration and communication of participants with ASD happened after practicing in the individual fine motor tasks. These results demonstrated that fine motor skill practice might have positive impacts on how participants with ASD are able to communicate and collaborate with partners in team-based activities.

Our findings underscore that social and communication skills do not occur in isolation; they relate not only to each other, but also to the complexity of the world around us. For people with ASD and fine motor delays, many important, everyday social tasks with motor components (such as playing a game) could be doubly affected, by ASD as well as these motor challenges, further straining EF skills and impacting performance. Motor skills are a common target of standardized intervention and assessment across the lifespan. For example, young children with ASD and fine motor delays often receive occupational therapy services embedded in their classroom special education supports [Baranek et al. 2008, Sowa and Meulenbroek 2012]. For older adolescents, emerging research supports the use of tasks such as collaborative Lego building [Wainer et al. 2010, Hsukens 2015] for assessment of work relevant skills. If combined, social and motor demands place increased strain on already stressed EF skills for individuals with ASD. We believe that a multicomponent approach to assessment and intervention may benefit many people.

Our system, to our knowledge, is the first to explore the influence of movement and grip control skills on the social communication and interaction skills of children with ASD through virtual environments. We provided collaborative fine motor tasks that require participants to plan, coordinate, and communicate about their motor manipulations. Additionally, our tasks provided haptic input which may make them more inherently rewarding and engaging. Our results suggested that improved collaborative task performance and increased active communication and collaboration were related to improved individual fine motor task performances. This supports the findings of MacDonald [MacDonald et al. 2013] who reported that children with weaker motor skills have greater social communicative skill deficits, and Srinivasan et al. [Srinivasan et al. 2015] who suggested socially embedded movement-based contexts are valuable in promoting imitation/praxis, interpersonal synchrony and motor performance. Our activities hold potential for supporting, intervening, and measuring multiple skills important for social success, paving the way for automated systems that minimize the need for intensive therapist oversight.

Though the presented work is promising, the results should be interpreted within the context of several limitations. First, the relatively small sample size and short intervention duration of the study, and the absence of a control group undermine the generalizability of the results. It is important and necessary to understand how such systems can be deployed over time, how skills transfer to other settings (e.g., in real-world settings), and how they can impact skills for children with ASD. Future studies will include more participants, training sessions and matched control groups to investigate generalization, different task modalities, therapist involvement, and use in community settings. Second, more virtual tasks and the randomized training mode should be developed, in order to reduce the learnability effect and maintain the users' interest. Third, though fine motor skills were shown to positively affect social interaction performance, more specific metrics should be developed to explore how they map to each other as well as EF domains. In addition, the human-human collaborative mode in the current system requires two users to align their availability. A stand-by partner (e.g., a virtual agent) would be preferable to perform the collaborative training at any time. This kind of agent system would also be helpful to reduce the burden for manually analyzing the communication data. Therefore, we will integrate a human-agent collaborative mode in C-Hg system in the future work. Despite these limitations, the C-Hg system is one of the first computer-assisted systems that enables simultaneous fine motor skill training and social skill training for children with ASD. The study results indicated that this system was acceptable among children and supported our hypothesis that improved collaborative task performance and increased active communication and collaboration were related to improved individual fine motor task performance. The encouraging results provide important preliminary insights into the development of more comprehensive multi-skill training environments.

References

[1] Hill, E.L., *Evaluating the theory of executive dysfunction in autism.* Developmental review, 2004. **24**(2): p. 189-233.

[2] Ozonoff, S. and J. Jensen, *Brief report: Specific executive function profiles in three neurodevelopmental disorders.* Journal of autism and developmental disorders, 1999. **29**(2): p. 171-177.

[3] Ozonoff, S., et al., *Executive function abilities in autism and Tourette syndrome: An information processing approach.* Journal of Child Psychology and Psychiatry, 1994. **35**(6): p. 1015-1032.

[4] McLean, R.L., et al., *Executive function in probands with autism with average IQ and their unaffected first-degree relatives.* Journal of the American Academy of Child & Adolescent Psychiatry, 2014. **53**(9): p. 1001-1009.

[5] Biederman, J., et al., *Impact of psychometrically defined deficits of executive functioning in adults with attention deficit hyperactivity disorder.* American Journal of Psychiatry, 2006. **163**(10): p. 1730-1738.

[6] Law, A.S., et al., *The impact of working memory load on task execution and online plan adjustment during multitasking in a virtual environment.* The Quarterly Journal of Experimental Psychology, 2013. **66**(6): p. 1241-1258.

7. Mackinlay, R., T. Charman, and A. Karmiloff-Smith, *High functioning children with autism spectrum disorder: A novel test of multitasking.* Brain and cognition, 2006. **61**(1): p. 14-24.

8. Rajendran, G., et al., *Investigating multitasking in high-functioning adolescents with autism spectrum disorders using the Virtual Errands Task.* Journal of autism and developmental disorders, 2011. **41**(11): p. 1445-1454.

9. White, S.J., P.W. Burgess, and E.L. Hill, *Impairments on "open-ended" executive function tests in autism.* Autism Research, 2009. **2**(3): p. 138-147.

10. Burgess, P.W., et al., *The cognitive and neuroanatomical correlates of multitasking.* Neuropsychologia, 2000. **38**(6): p. 848-863.

11. McGeorge, P., et al., *Using virtual environments in the assessment of executive dysfunction.* Presence: Teleoperators & Virtual Environments, 2001. **10**(4): p. 375-383.

12. Hutchison, S.M., U. Müller, and G. Iarocci, *Parent Reports of Executive Function Associated with Functional Communication and Conversational Skills Among School Age Children With and Without Autism Spectrum Disorder.* Journal of autism and developmental disorders, 2019: p. 1-11.

13. Rao, P.A., D.C. Beidel, and M.J. Murray, *Social skills interventions for children with Asperger's syndrome or high-functioning autism: A review and recommendations.* Journal of autism and developmental disorders, 2008. **38**(2): p. 353-361.

14. Gowen, E. and A. Hamilton, *Motor abilities in autism: a review using a computational context.* Journal of autism and developmental disorders, 2013. **43**(2): p. 323-344.

15. Ming, X., M. Brimacombe, and G.C. Wagner, *Prevalence of motor impairment in autism spectrum disorders.* Brain and Development, 2007. **29**(9): p. 565-570.

16. Jansiewicz, E.M., et al., *Motor signs distinguish children with high functioning autism and Asperger's syndrome from controls.* Journal of autism and developmental disorders, 2006. **36**(5): p. 613-621.

17. Forti, S., et al., *Motor planning and control in autism. A kinematic analysis of preschool children.* Research in Autism Spectrum Disorders, 2011. **5**(2): p. 834-842.

18. MacDonald, M., C. Lord, and D.A. Ulrich, *The relationship of motor skills and social communicative skills in school-aged children with autism spectrum disorder.* Adapted Physical Activity Quarterly, 2013. **30**(3): p. 271-282.

19. David, F.J., et al., *Coordination of precision grip in 2-6 years-old children with autism spectrum disorders compared to children developing typically and children with developmental disabilities.* 2012.

20. Crippa, A., et al., *Eye-hand coordination in children with high functioning autism and Asperger's disorder using a gap-overlap paradigm.* Journal of autism and developmental disorders, 2013. **43**(4): p. 841-850.

21. Johnson, B.P., et al., *A quantitative comparison of handwriting in children with high-functioning autism and attention deficit hyperactivity disorder.* Research in Autism Spectrum Disorders, 2013. **7**(12): p. 1638-1646.

22. Chetcuti, L., et al., *Object-directed imitation in autism spectrum disorder is differentially influenced by motoric task complexity, but not social contextual cues.* Autism, 2019. **23**(1): p. 199-211.

23. Srinivasan, S.M., et al., *The effects of rhythm and robotic interventions on the imitation/praxis, interpersonal synchrony, and motor performance of children with autism spectrum disorder (ASD): a pilot randomized controlled trial.* Autism research and treatment, 2015. **2015**.

24. Fulceri, F., et al., *Motor skills as moderators of core symptoms in Autism Spectrum Disorders: preliminary data from an exploratory analysis with Artificial Neural Networks.* Frontiers in psychology, 2018. **9**: p. 2683.

25. Zhao, H., et al., *Hand-in-hand: A communication-enhancement collaborative virtual reality system for promoting social interaction in children with autism spectrum disorders.* IEEE Transactions on Human-Machine Systems, 2018. **48**(2): p. 136-148.

26. Zheng, Z., et al., *Design, development, and evaluation of a noninvasive autonomous robot-mediated joint attention intervention system for young children with ASD.* IEEE Transactions on Human-Machine Systems, 2017. **48**(2): p. 125-135.

27. Battocchi, A., et al. *Collaborative Puzzle Game: a tabletop interactive game for fostering collaboration in children with Autism Spectrum Disorders (ASD)*. in *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. 2009. ACM.

28. Sallnäs, E.-L. *Haptic feedback increases perceived social presence*. in *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. 2010. Springer.

29. Basdogan, C., et al., *An experimental study on the role of touch in shared virtual environments.* ACM Transactions on Computer-Human Interaction (TOCHI), 2000. **7**(4): p. 443-460.

30. Rosenblum, S., H.A.B. Simhon, and E. Gal, *Unique handwriting performance characteristics of children with high-functioning autism spectrum disorder.* Research in Autism Spectrum Disorders, 2016. **23**: p. 235-244.

31. Palsbo, S.E. and P. Hood-Szivek, *Effect of robotic-assisted three-dimensional repetitive motion to improve hand motor function and control in children with handwriting deficits: a nonrandomized phase 2 device trial.* American Journal of Occupational Therapy, 2012. **66**(6): p. 682-690.

32. Kim, Y.-S., et al. *Haptics Assisted Training (HAT) System for children's handwriting*. in *World Haptics Conference (WHC), 2013*. 2013. IEEE.

33. Zhao, H., et al. *Understanding Fine Motor Patterns in Children with Autism Using a Haptic-Gripper Virtual Reality System*. in *International Conference on Universal Access in Human-Computer Interaction*. 2018. Springer.

34. Zhao, H., et al., *Design of a Haptic-Gripper Virtual Reality System (Hg) for Analyzing Fine Motor Behaviors in Children with Autism.* ACM Transactions on Accessible Computing (TACCESS), 2018. **11**(4): p. 19.

35. *The Geomagic Touch Haptic Device*. Available from: http://www.geomagic.com/en/products/phantom-omni/overview.

36. Colgate, J.E., M.C. Stanley, and J.M. Brown. *Issues in the haptic display of tool use*. in *Intelligent Robots and Systems 95.'Human Robot Interaction and Cooperative Robots', Proceedings. 1995 IEEE/RSJ International Conference on*. 1995. IEEE.

37. Adams, R.J. and B. Hannaford, *Stable haptic interaction with virtual environments.* IEEE Transactions on robotics and Automation, 1999. **15**(3): p. 465-474.

38. Sankaranarayanan, G. and B. Hannaford. *Virtual coupling schemes for position coherency in networked haptic environments*. in *Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on*. 2006. IEEE.

39. Niemeyer, G. and J.-J.E. Slotine, *Telemanipulation with time delays.* The International Journal of Robotics Research, 2004. **23**(9): p. 873-890.

40. Niemeyer, G. and J.-J. Slotine, *Stable adaptive teleoperation.* IEEE Journal of oceanic engineering, 1991. **16**(1): p. 152-162.

41. Yang, C., et al., *Teleoperation control based on combination of wave variable and neural networks.* IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2017. **47**(8): p. 2125-2136.

42. Wang, Z., et al., *Individuals with autism spectrum disorder show abnormalities during initial and subsequent phases of precision gripping.* Journal of neurophysiology, 2015. **113**(7): p. 1989-2001.

43. Anzulewicz, A., K. Sobota, and J.T. Delafield-Butt, *Toward the Autism Motor Signature: Gesture patterns during smart tablet gameplay identify children with autism.* Scientific reports, 2016. **6**: p. 31107.

44. Cook, J.L., S.-J. Blakemore, and C. Press, *Atypical basic movement kinematics in autism spectrum conditions.* Brain, 2013. **136**(9): p. 2816-2824.

45. C. Lord et al., *The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism*, (in English), Journal of Autism and Developmental Disorders, vol. 30, no. 3, pp. 205-223, Jun 2000.

46. Constantino, J.N. and C.P. Gruber, *Social responsiveness scale (SRS)*. 2007: Western Psychological Services Los Angeles, CA.

47. Rutter, M., A. Bailey, and C. Lord, *The social communication questionnaire: Manual*. 2003: Western Psychological Services.

48. Keith E Beery, N.A.B., Natasha Beery, *The Beery-Buktenica Developmental Test of Visual-Motor Integration: Beery VMI with Supplemental Developmental Tests of Visual Perception and Motor Coordination: Administration, Scoring and Teaching Manual (6th ed.)*. 2010, Minneapolis, MN: NSC Pearson.

49. Wilcoxon, F., *Individual comparisons by ranking methods.* Biometrics bulletin, 1945. **1**(6): p. 80-83.

50. Cohen, J., *Statistical power analysis for the behavioral sciences*. 1988, Hillsdale, N.J.: L. Erlbaum Associates.

51. Baranek, G.T., L. Wakeford, and F. David, *Understanding, assessing, and treating sensory–motor issues.* Autism spectrum disorders in infants and toddlers: Diagnosis, assessment, and treatment, 2008: p. 104-140.

52. Sowa, M. and R. Meulenbroek, *Effects of physical exercise on autism spectrum disorders: a meta-analysis.* Research in Autism Spectrum Disorders, 2012. **6**(1): p. 46-57.

53. Wainer, J., et al., *The effectiveness of using a robotics class to foster collaboration among groups of children with autism in an exploratory study.* Personal and Ubiquitous Computing, 2010. **14**(5): p. 445-455.

54. Huskens, B., et al., *Improving collaborative play between children with autism spectrum disorders and their siblings: The effectiveness of a robot-mediated intervention based on Lego® therapy.* Journal of autism and developmental disorders, 2015. **45**(11): p. 3746-3755.

55. Zhao, H (2021). *Modeling and Analysis of Multimodal Collaborative Virtual Interaction for Autism Spectrum Disorder Intervention* [Unpublished doctoral thesis]. Vanderbilt University

56. Zhang, L., Warren, Z., Swanson, A., Weitlauf, A. and Sarkar, N., 2018. Understanding performance and verbal-communication of children with ASD in a collaborative virtual environment. Journal of autism and developmental disorders, 48(8), pp.2779-2789.

57. Owen-DeSchryver, J. S., Carr, E. G., Cale, S. I., & Blakeley-Smith, A. (2008). Promoting social interactions between students with autism spectrum disorders and their peers in inclusive school settings. Focus on Autism and Other Developmental Disabilities, 23(1), 15–28.

This chapter is reprinted with permission from Huan Zhao, Ashwaq Zaini Amat, Miroslava Migovich, Amy Swanson, Amy S. Weitlauf, Zachary Warren, and Nilanjan Sarkar. 2021. C-Hg: A Collaborative Haptic-Gripper Fine Motor Skill Training System for Children with Autism Spectrum Disorder. ACM Trans. Access. Comput. 14, 2, Article 9 (June 2021), 28 pages. https://doi.org/10.1145/3459608

CHAPTER 4: DESIGN OF AN INTERACTIVE VIRTUAL REALITY SYSTEM, INVIRS, FOR JOINT ATTENTION PRACTICE IN AUTISTIC CHILDREN

4.1 Abstract

Many children with Autism Spectrum Disorder (ASD) exhibit atypical gaze behaviors related to joint attention, a fundamental social-communication skill. Specifically, children with ASD1 show differences in the skills of gaze sharing and gaze following. In this work we present a novel virtual reality (VR)-based system, called InViRS, in which children with ASD play games allowing them to practice gaze sharing and gaze following. InViRS has three main design contributions: (i) a closed-loop joint attention paradigm with real-time tracking of the participant's eye gaze and game performance measures, (ii) an assistive feedback mechanism that provides guidance and hints in real time, and (iii) a controller that adaptively changes the avatar's gaze prompts according to the performance measures. Results from a pilot study to evaluate the feasibility of InViRS with 9 autistic1 children and 9 typically developing (TD) children offered preliminary support for the feasibility of successful gameplay as well as positive impacts on the targeted skills of gaze sharing and gaze following.

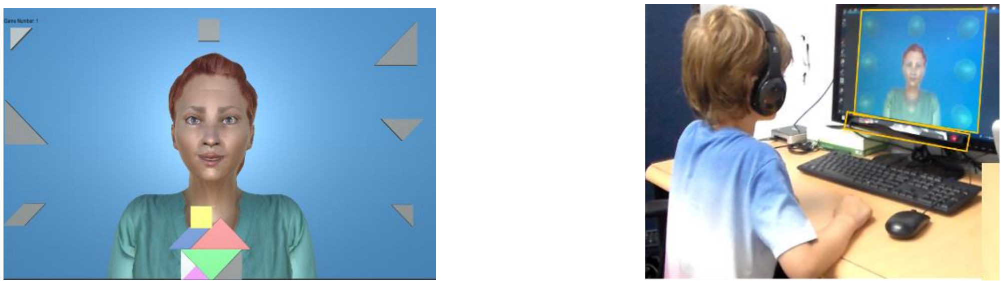
4.2 Introduction

Autism spectrum disorder (ASD) affects approximately 1 in 54 children in the US [1] with significant associated costs [2]. Many children with ASD experience impairment in joint attention – a fundamental social skill that requires gaze sharing and gaze following with another person. Joint attention, which is different from simply making eye contact, is crucial to learning new information, knowledge exchange, and early language development [3 - 5]. Joint attention skills can be defined as the ability to coordinate one's attention with another person towards an object or an event of interest [6]. There are two main components in joint attention: gaze sharing and gaze following. In gaze sharing, one is required to be aware of the other person's gaze and intent to share information. In gaze following, which emerges after gaze sharing, one is required to shift one's gaze and attention to the object or event being shared. Joint attention can be initiated by another person, which is known as response to joint attention (RJA) or can be initiated on their own, which is known as initiation of joint attention (IJA). Behavior-based interventions have shown promise in imparting joint attention skills in young children [7, 8], but their cost and trained personnel requirements limit their availability [7].

Although not posited as a replacement for skilled clinical care, technology-based interventions can complement and support behavioral intervention by increasing attention and learning in autistic individuals [9], many of whom show an affinity for technology [10]. Virtual reality (VR) based intervention, although not a substitute for human intervention, can provide a safe environment wherein autistic children can interact with a system to practice their skills [11]. To assess engagement and response, VR can be integrated with peripheral sensors such as eye trackers and physiology sensors to provide measures of eye gaze [12] and physiological response [13, 14]. In recent years, VR-based joint attention studies have explored gaze perception, cognition, focus, and engagement in autistic individuals during joint attention interaction [15 - 18]. However, only a few studies [15, 16] have examined gaze sharing and gaze following specifically.

The primary contribution of this work is the design, development, and preliminary assessment of a novel Interactive Virtual Reality System (InViRS), an adaptive game-based system for practicing core joint attention skills of gaze sharing and gaze following. In InViRS, a RJA paradigm initiated by a virtual avatar acts as an interaction partner that provides participants with gaze prompts through a closed-loop joint attention paradigm and real-time hints using continuous measurement of eye gaze and game performance. Rather than attempting to train individuals to make sustained eye contact, which many people with ASD describe as uncomfortable [19, 20], this system instead teaches them how to use another person's gaze to gather important information about the environment as well as that person's intentions and interests.

The current work substantially expands our previous conference paper [21] in terms of i) system augmentation, ii) introduction of an individualized adaptation model and iii) data from a pilot study. System augmentation included adding a new dimension to the avatar's gaze prompts by manipulating the depth of the eye movements together with varying speed of the avatar's gaze prompts and the inclusion of new region of interests on the avatar's face to observe participants' gaze fixation in a detailed manner. In addition, we present new results of a pilot study involving autistic and typically developing (TD) children.



(a)

(b)

Figure 4-1: The virtual game environment. (a) Tangram Puzzle game. (b) A participant playing the Bubble Popping game.

The presented research contributes to the design of a real-time gaze detection algorithm, a task difficulty adjustment algorithm, an avatar controller that adjusts the avatar's behaviors, and a supervisory controller that has embedded logic to coordinate the closed-loop interaction for individualized joint attention practice based on real-time measurement. Such a system itself is novel in this field and in our opinion, contributes towards the design of a new adaptive behavioral intervention system for ASD. Endowing InViRS with these abilities allows us to analyze RJA performance at the component level - gaze sharing and gaze following performances - in addition to overall RJA performance, a uniquely important contribution to this area of research, as the technologically facilitated ability to parse joint attention skills at a more granular level will potentially allow the development of targeted behavioral intervention. The remainder of the paper is organized as follows: Section 2 presents relevant literature reviews; Section 3 describes system design and architecture; and Sections 4 and 5 present the experimental setup and the results of the study, respectively. Finally, Section 6 presents discussion on the potential and limitations of the current study.

4.3 System Design

InViRS was developed as a game-based system through which children with ASD can practice the skills of gaze sharing and gaze following. Although InViRS is capable of delivering multiple game modes, in its current form, children play two different games with a virtual avatar: a Tangram Puzzle game, used for practice, and a Bubble Popping game, used for pre- and post-assessment (see Figures 4-1(a) and 4-1(b), respectively, and section 4.3.1). Research shows that simple puzzle games are engaging for children with ASD [49]. We chose the Tangram puzzle game for joint attention practice in the hope that it would keep participants engaged. It was not too complex so as not to frustrate the participants, but at the same time had enough variation to keep the participants interested. We also wanted to choose a simple game for pre and post assessment that was both easy to control and visually interesting. The Bubble Popping game satisfied both these criteria. Both games were successfully used in our previous work with children with ASD [50, 51]. Each game involves systematic assessment of children's eye gaze in response to scaffolded prompts, across varying difficulty levels. InViRS has several options to create individualized and adaptive interaction with the child: 1) provision of varying gaze prompts, 2) delivery of prompts and visual aids using the least-to-most (LTM) prompting mechanism, 3) an adaptive module that changes the avatar's interaction level to match the participant's performance, 4) variation in the speed of gaze prompts to actively probe participant's ability to follow gaze, and 5) real-time computation of game performance.

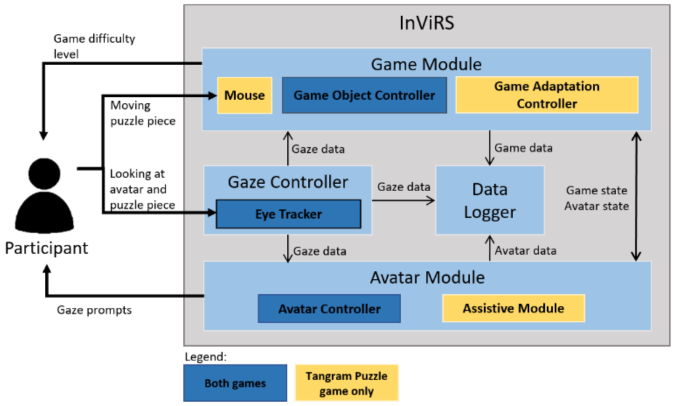


Figure 4-2: Human-Computer Interaction block diagrams for InViRS. The Game Adaptation Controller and the Assistive Module are not activated for the Bubble Popping game.

4.3.1 InViRS Games and Human-Computer Interaction

Figure 4-2 illustrates the interaction diagrams between the participant and InViRS. The eye tracker and mouse captured the participant's gaze data in both games and puzzle pieces movement in the Tangram Puzzle game. The Gaze Controller i) sends gaze data to the Avatar Module to trigger the avatar's gaze prompts, ii) updates the Game Module, and iii) logs the gaze data in the Data Logger. The Game Module manages the difficulty level of the game through the Game Adaptation Controller where difficulty level can be changed based on the gaze data, game states, and avatar states. The Assistive Module in the Avatar Module provides hints and assistance based on the participant's performance.

Note that because of the structure of the Bubble Popping game, only the eye gaze data from the eye tracker are used to interact with the avatar and select the correct bubble to pop. Since there is no Assistive Module or Game Adaptation Controller in this game, the avatar's gaze prompts and game difficulty level are increased continuously without any assistance or adaptive adjustments to the difficulty level.

4.3.1.1 Gaze Sharing

Within InViRS, gaze sharing is defined when a participant fixates their gaze on a predefined region around the avatar's eye (Figure 4-3), and not necessarily directly on the avatar's eyes. This was designed so that gaze sharing could be established without inducing the stress that may be evoked within individuals with ASD when they are forced to make direct, sustained eye contact [18, 19]. We chose a minimum duration for fixation of 200 ms based on the study presented by Rayner as a reasonable human gaze fixation characteristic [41]. When a gaze lasts more than 200 ms , the avatar will trigger the next prompt by shifting its gaze towards a game object (either at a puzzle piece in the Tangram Puzzle game or at a bubble in the Bubble Popping game).

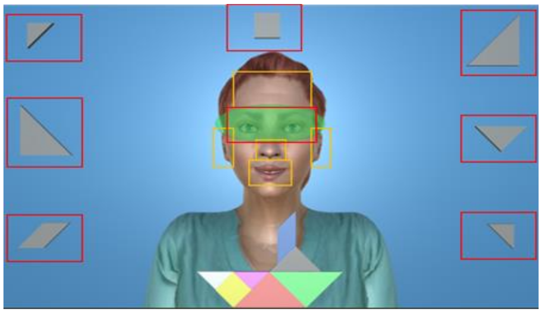


Figure 4-3: The ROIs for the Tangram Puzzle game. Red boxes represent active ROIs and yellow boxes represent passive ROIs.

We setup InViRS to wait for 30 seconds for a gaze to be registered on the avatar's eye region before progressing to the next state. We chose 30 seconds in consultation with clinical psychologists specializing in ASD intervention as we wanted to give enough time for the children to receive the cue, process and respond to the avatar's prompt.

Longer waiting time might cause the children to lose focus and interest in the game. If participants did not look at the avatar's eye region within 30 seconds, the system provided audio and visual cues. In the Tangram Puzzle game (practice), an audio cue in the form of 3 seconds of bell ringing was played and a visual cue of highlighting the avatar's eye region was provided. In the Bubble Popping game (assessment), only the 3 seconds bell ringing audio cue was played if participants did not look. For both games, if no eye contact was made within 2 minutes, the game was terminated.

4.3.1.2 Gaze Following

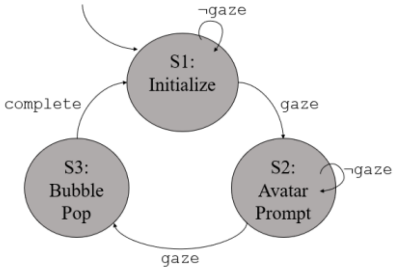
As mentioned previously, after a participant successfully share their gaze with the avatar, InViRS triggers an event for the avatar to direct its gaze at a game object. The participant then needed to direct their gaze to the game object that was prompted to trigger the next event in InViRS.

In the Tangram Puzzle game, after the participant looked at the correct game object, the color of the object was revealed and the participant could move the puzzle piece to the target area using the mouse. If a participant did not look at the correct game object within 30 seconds, InViRS triggered assistive events from the Assistive Module to get the participant to look at the intended area. For example, the avatar would repeat the gaze prompt at a slower pace together with highlighting the puzzle piece it prompted. Details of the assistance for the Tangram Puzzle game is presented in 4.3.6.

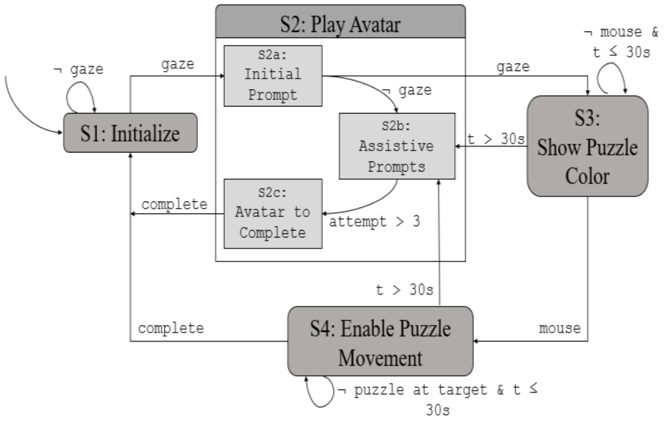
As for the Bubble Popping game, when the participant looked at the correct bubble, the bubble would pop and new bubbles will be generated. If no gaze was detected on the correct bubble within 30 seconds, no assistive events were triggered and the avatar proceeded to provide the next gaze prompt.

4.3.2 Virtual Game Environment

The virtual game environment was developed using Unity v5.6.1f1 [22], a widely utilized virtual game development tool. Both games in the virtual environment were developed as finite state machines (FSM). We defined a 5-tuple deterministic FSM as detailed in Table 4-1. Figure 4-4 illustrates the FSMs for both games.



(a)



(b)

Figure 4-4: Finite state machines (FSM) for InViRS virtual environment. (a) FSM for the Bubble Popping game. (b) FSM for the Tangram Puzzle game

Table 4-1: FSM Tuple

Tuple	Definition	Bubble Popping game	Tangram Puzzle game
Q	set of states	{Initialize, Avatar Prompt, Bubble Pop}	{Initialize, Play Avatar, Show Puzzle Color, Enable Puzzle Movement}
Σ	set of inputs	{gaze, complete}	{gaze, mouse, complete}
q_0	initial state	Initialize	Initialize
F	set of final states	Initialize	Initialize

4.3.3 Gaze Controller

In this study, we designed a controller that used eye tracking data from a Tobii EyeX [23] eye tracker in real-time to perform gaze analysis. The sampling frequency of the eye tracker is comparatively low, between 50-60 Hz, but is sufficient for use in this study, as the primary interest is on fixation data points rather than pupil diameter, saccades, and other fast-moving gaze points [24]. We used a Tobii-Unity development package [25] to: i) continuously collect gaze points during game play, and ii) register a gaze fixation on a predefined region when a gaze duration of approximately 200 ms [41] was measured. The gaze points that were collected in this controller were sent to the Data Logger to be recorded together with the time stamp and game state at that time.

Additionally, we defined several regions of interest (ROIs) in Unity to capture participant's gaze on these areas. There were two categories of ROIs, active and passive, created for the objects and avatar in the games. The active ROIs were defined on the avatar's eye region and all game objects in the games (puzzle pieces and bubbles). Taking into consideration the difficulty in autistic children to look directly at someone's eye gaze [19, 20], we defined a rectangular region around the avatar's eye to reduce discomfort when establishing gaze sharing. When a gaze was first detected on the avatar's eye ROI, the controller would start a timer to measure the duration of the gaze. If the duration was more than 200 ms [41], the controller would trigger an event to the Avatar Module to indicate gaze sharing was initiated. If the duration of the gaze was less than 200 ms [41], the gaze would not trigger any event and the timer was reset before a new gaze was detected on the eye region again. The same algorithm was used when a gaze was detected on a game object ROI. If the gaze was detected on the correct game object for 200 ms, the controller would trigger an event to the Game Module to indicate that the correct game object was looked at.

As for the passive ROIs, five facial areas of the avatar were selected that included: the forehead, right ear, left ear, nose, and mouth. When a gaze was detected on a passive ROI, the controller would send the name, location and time stamp of the ROI to the Data Logger to be recorded. Figure 4-3 shows all the ROIs in the Tangram Puzzle game environment. The ROIs definitions are not limited to the objects in the Tangram Puzzle and Bubble Popping games and can be used in other VR environments that focus on gaze analysis or where non-verbal interaction is of interest.

4.3.4 Avatar Controller

The design and animation of the avatar were accomplished using a 3D graphics application called Autodesk Maya [26]. The neutral facial expression for the avatar in this study was by design. Because the objective of this study was to evaluate the impact of a novel interactive virtual system on gaze sharing and gaze following, we chose a neutral expression to observe how participants responded to the eye gaze prompts without other factors, such as emotional valence, influencing the result. We customized the avatar’s head and eye movement such that the avatar could gaze in any direction to locate the relevant objects of the game. In this work, we created eight different gaze directions to correspond to the eight bubble pieces and seven tangram puzzle pieces. We also added different gaze prompt configurations for each gaze direction that consisted of animating the avatar’s head movement together with the eye movement, and manipulating the range of the movement of avatar’s eyeball from the center of the eye. Head movement has been shown to influence gaze following [27 - 29] eliciting faster response time when head and eye move congruently [30, 31]. As such, we used the head and eye movement together as the initial gaze prompts to represent an easy level. For the next gaze prompt difficulty level, we removed the head movement and only maintained the eye movements for gaze prompts. In this level, we had the avatar’s eye move from the center of the eye to the edge of the eye in the direction of the gaze prompt to represent maximum range of human eyeball movement [47]. For the third gaze prompt difficulty level, the avatar’s eyeball movement was reduced to 40% of the maximum movement range to create a subtle gaze prompt as judged by consensus of human observers. Figure 4-5 provides an example of the three gaze variations in the upper right direction. The combination of using gaze prompts in varying direction, depth of eye movement and speed in this study demonstrates the flexibility of our avatar’s design that can be easily configured to support other gaze related implementations.

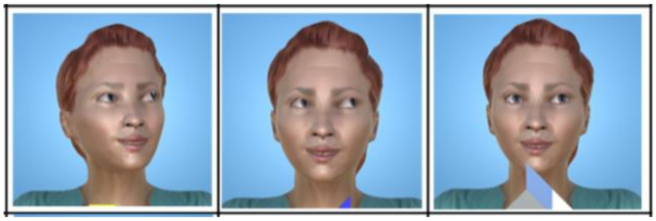


Figure 4-5: Example of avatar's different eye gaze configurations in upward right direction. (a) Head movement together with eye movement, (b) Full eye movement, and (c) Minimal eye movement

In both games, the gaze directions were randomly selected to avoid predictive behavior. For the Tangram Puzzle game, the different gaze prompt levels were evenly implemented as described in Table 4-2. As for the Bubble Popping game, the gaze prompt level was kept at the second difficulty level and only the speed of the prompts was continuously increased.

Table 4-2: Different levels of avatar gaze prompts

Game Number	Avatar gaze prompt level
1, 2, 3	Head movement + Full eye movement
4, 5, 6	Full eye movement
7, 8, 9	Minimal eye movement

4.3.5 Game Adaptation Controller

The Game Adaption Controller is a part of Tangram Puzzle game that managed the change in the avatar's interaction level with the participant based on participant's performance. A rule-based adaptive algorithm was developed by using both game performance and gaze data as inputs to change i) the avatar's gaze prompt level (as per Table 4-2) and ii) the speed of the avatar's gaze prompts. In addition to varying the avatar's gaze prompt level, we also changed the speed of the avatar's gaze prompts to make the game more challenging. The higher the speed of the gaze prompt, the harder it was for the participant to follow the gaze. For the Bubble Popping game, we did not use the Game Adaptation Module. The speed of the avatar's gaze prompt in that game was increased at a constant rate in each prompt regardless of the participant's performance in the Bubble Popping game.

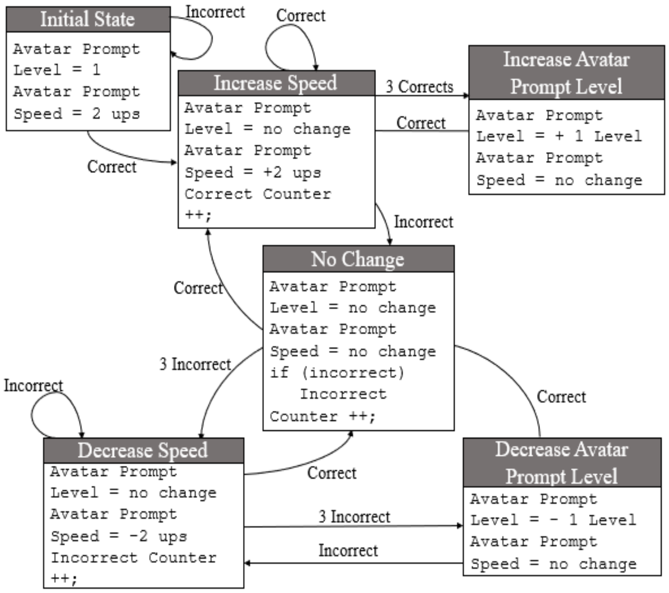

Initial State
Avatar Prompt Level = 1 Avatar Prompt Speed = 2 ups
Incorrect
Correct
Increase Speed
Avatar Prompt Level = no change Avatar Prompt Speed = +2 ups Correct Counter ++;
3 Corrects
Correct
Increase Avatar Prompt Level
Avatar Prompt Level = + 1 Level Avatar Prompt Speed = no change
Correct
Incorrect
No Change
Avatar Prompt Level = no change Avatar Prompt Speed = no change if (incorrect) Incorrect Counter ++;
Correct
3 Incorrect
Incorrect
Correct
Decrease Speed
Avatar Prompt Level = no change Avatar Prompt Speed = -2 ups Incorrect Counter ++;
Correct
3 Incorrect
Incorrect
Decrease Avatar Prompt Level
Avatar Prompt Level = - 1 Level Avatar Prompt Speed = no change

Figure 4-6: State Diagram for Game Adaptation Controller for Tangram Puzzle game.

Figure 4-6 summarizes the adaptive algorithm. At the beginning of a Tangram Puzzle game, the gaze prompt level was set to Level 1 where the gaze prompt included the head movement together with eye gaze, while the speed of the avatar's gaze prompt was set to a rate of 2 units per second (ups). When a participant correctly chose a puzzle piece that was prompted by the avatar, the subsequent speed of the avatar's gaze prompt was increased at a constant rate of 2 ups. The speed remained the same when the participant failed to choose the correct puzzle piece. After three consecutive puzzle pieces were correctly selected, the gaze prompt level was increased such that the avatar's gaze prompt was reduced to only eye gaze movements. Whereas, after three consecutive wrong attempts of choosing the corresponding puzzle pieces, the speed of the next gaze prompt was reduced by 2 ups. Then, if the participant continues to make three more consecutive incorrect selections, the avatar's gaze prompt level was decreased to make the gaze prompts easier for the participant to follow and to provide opportunities for the participant to continuously strive and challenge their gaze following skills.

4.3.6 Assistive Module

The Assistive Avatar Module was used only in the Tangram Puzzle game to assist the participants when they were unable to direct their gaze at the correct ROIs or in the intended direction. This module was not used in the Bubble Popping game.

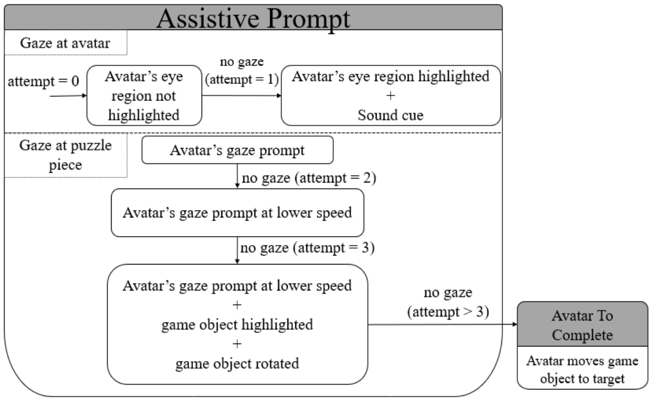

Assistive Prompt
Gaze at avatar
attempt = 0
Avatar's eye region not highlighted
no gaze (attempt = 1)
Avatar's eye region highlighted + Sound cue
Gaze at puzzle piece
Avatar's gaze prompt
no gaze (attempt = 2)
Avatar's gaze prompt at lower speed
no gaze (attempt = 3)
Avatar's gaze prompt at lower speed + game object highlighted + game object rotated
no gaze (attempt > 3)
Avatar To Complete
Avatar moves game object to target

Figure 4-7: Flow chart of the avatar's Assistive Prompt. Number of attempts increased when participant was unable to look at the correct place or game object.

The assistive avatar module used a least-to-most (LTM) prompting mechanism [32], which is widely used in intervention for children with ASD. The principle of LTM is to allow the learner the opportunity to independently execute the task with the least amount of prompting, which is then increased progressively depending on the need. The LTM mechanism has also been previously used to teach communication skills [33 - 35], and motor skills [36] in children with ASD. In this current study, LTM implies allowing the participant to interpret the avatar's gaze prompt on their own before the avatar provides additional prompts leading the participant to the correct game object.

Within our LTM design, we used both real-time gaze and current performance data as inputs to create a personalized assistance to the participants. For example, a participant performing at a higher gaze prompt level and higher prompt speed will receive a different assistive prompt compared to a participant performing at a lower gaze prompt level or prompt speed. This module supports individualized learning conditions across different participants' performance levels. Figure 4-7 shows the progression of the assistive prompts for every unsuccessful attempt and Table 4-3 lists the assistance the avatar provided in order of the number of attempts the participant made.

4.3.7 Game Object Controller

The Game Object Controller manages the configuration of the game objects in both games. In the Bubble Popping game, this controller initialized the bubbles into their respective location in the virtual space. When a gaze event on the target bubble was received from the Gaze Controller, the Game Object Controller enabled the bubble to pop and waited 5 seconds before the bubble was regenerated at the same original location again. As for the Tangram Puzzle game, the controller initialized the puzzle pieces to their initial locations, set the appearance of each puzzle piece to zero color saturation (grayscale), and disabled their movements. When a gaze event on the target piece was received from the Gaze Controller, the Game Object Controller: i) displayed the color of the puzzle piece, ii) enabled movement of the puzzle piece, and

iii) updated the movement of the puzzle piece to the target location. Once all the puzzle pieces were at the target location, the controller triggered an event to the game settings component to indicate the completion of the game and proceeded to the next game. This controller also tracks other game properties including the number of games played, duration of each game, points accumulated, and the number of assistances a participant used in each move.

Table 4-3: Assistive prompts in Tangram Puzzle game

No. of Attempts	Assistive Prompts	Reason for assistance
0	(1) Highlight avatar's eye region	Initial condition
1	(1) Highlight avatar's eye region + (2) Sound cue	Participant did not make eye contact with the avatar
2	(1) Avatar repeats gaze prompt at a lower speed	Participant did not select the correct game object
3	(1) Avatar repeats gaze prompt at a lower speed + (2) Highlight the game object + (3) Rotating game object in place	Participant did not select the correct game object
>3	(1) Avatar automatically moves the game object to the target location	Participant did not select the correct game object

4.3.8 Data Logger

The data logger collected all the virtual environment data for real-time manipulation in the adaptive module and for offline data analysis. The real-time data used by the adaptive algorithm included participant's game score, gaze ROIs, and avatar configurations.

4.4 Experimental Design

We conducted a pilot study to evaluate the hypotheses that practicing in InViRS would be able to: i) improve gaze sharing in autistic children as indicated by increased in fixation frequency and duration on the eye region but not necessarily directly on the eye as compared to other facial features during interaction, and ii) improve gaze following skills in autistic children represented by improved game score. Additionally, we also wanted to compare game and gaze performance between ASD and TD participants to identify any meaningful differences. We administered a pre-test and post-test to assess changes in gaze fixation, gaze following, and performance measures after participating in practice session.

4.4.1 Participants

We recruited a total of 18 children (9 children with ASD, 9 TD children) to participate in the study. The age range of the participants was between 7 and 13 years. Children with ASD were recruited from a large research registry maintained by the Vanderbilt Kennedy Center of children previously diagnosed with ASD by licensed clinical psychologists using standard diagnostic tools, such as the Autism Diagnostic Observation Schedule (ADOS) [37]. The TD children were recruited from the local community through regional advertisement.

To assess the current level of ASD symptoms of all participants and ensure baseline symptom differences between diagnostic groups, parents of all participants were asked to fill out the Social Communication Questionnaire (SCQ) [38] and the Social Responsiveness Scale, Second Edition (SRS-2) [39]. Both scales provide quantitative measures of observable characteristics of ASD via paper-and-pencil parent report. In this study, we used the SCQ Lifetime Total Score. This score ranges from 0 to 39, with a score above 15 indicative of likely ASD. For the SRS-2, participants received a Total Score and a T-score. A Total Score of 98 or a T-score value of 76 reflects high risk of ASD. Table 4-4 presents the characteristics of the participants.

This study was approved by the Institutional Review Board at Vanderbilt University (IRB Number: 180047). Consents from the participants' guardians and assents from the participants themselves were obtained before the experiment were conducted. A gift card was presented to participants at the conclusion of each visit.

Table 4-4: Characteristics of Participants

Participants	ASD (n = 9)	TD (n = 9)
	Mean (SD)	Mean (SD)
Age	11.00 (1.35)	10.98 (1.98)
Gender (% male)	55.6 %	55.6 %
SCQ Lifetime Total Score	21.56 (7.33)	2.33 (2.69)
SRS-2 Total Score	101.78 (18.54)	24.00 (27.06)
SRS-2 T-score	78.22 (7.38)	48.44 (16.12)

SRS-2: Social Responsiveness Scale, Second Edition
SCQ: Social Communication Questionnaire

4.4.2 Protocol

The study consisted of three visits with 5 to 10 days between visits. In the first visit, the participants completed a pre-test which was the Bubble Popping game before starting the Tangram Puzzle practice game, and at the last visit, they completed another Bubble Popping game for post-test after finishing the last practice Tangram Puzzle game. The second visit was fully dedicated to practice with the Tangram Game. The order of each game was important since we needed to make sure that practice games were administered between the pre-test and post-test. At each visit, before starting any games, a participant's eye gaze was calibrated on the Tobii EyeX eye tracker.

4.5 Results

Five performance metrics were defined to evaluate the hypotheses stated in Section III based on the results obtained from the Bubble Popping game in the pre- and post-tests. Table 4-5 lists the metrics together with a description of each metric. All statistical analyses were performed using MATLAB statistical computation functions. In this study, we calculated gaze fixation points in MATLAB using the EyeMMV toolkit [40].

Table 4-5: List of Performance Metrics

Performance Metric	Description
Score	One point is received when a participant looked at the correct game object (i.e., a target bubble) that was prompted by the avatar. Maximum possible score is 50.
Time to complete (seconds)	Total time it takes by a participant to interact with the avatar and selecting the bubble for all 50 gaze prompts. Game is terminated if 120 seconds pass by without any interaction by the participant at all.
Response time (seconds)	Response time is computed between the time when the avatar provides a gaze prompt and the time the participant looks at the correct bubble. The time is reset when no gaze interaction is detected after 30 seconds. After that time, the avatar provides a new gaze prompt and the timer starts again.
Fixation points	Gaze fixation was calculated using EyeMMV toolkit [40] in MATLAB based on ROIs parameters; i) name of the ROIs and ii) duration of gaze on ROIs. (Figure 5-4 illustrates all the facial ROIs)
Ratio of gaze fixation on eye to gaze fixation on other facial features	Ratio of number of gaze fixation points on the avatar's eye region compared to number of gaze fixation points on other facial ROIs

4.5.1 Overall Game Performance Measures

Game performance was measured using game score, time to complete the game, and the response time to each gaze prompt. First, on average, the autistic children improved their scores by 8 points in the post-test, which was closer to TD children's game score in the pre-test. However, this improvement was not statistically significant. Meanwhile, the TD children did not show much improvement in the post-test compared to the pre-test, which may indicate that the TD children were already performing at their highest level in the pre-test because the game was not difficult for them. Next, we found statistically significant improvement in the time to complete the Bubble Popping game measure for autistic children ($p = 0.0106$). They improved on average by 1 minute and 20 seconds in the post-test, while the TD children spent 23 seconds less on average in the post-test. Lastly, autistic children showed improvement in the time to respond to the avatar's gaze prompts measure, but the improvement was not statistically significant. On average they took 3.4 seconds to respond to the avatar's gaze prompt in the pre-test, while in the post-test, they took on average 1.7 seconds to respond. Meanwhile, TD children spent almost the same time to respond in both pre-test and post-test, which were 1.6 seconds and 1.2 seconds, respectively. When looking at the effect size of the ASD participants, we observed a large effect size for the time to complete category, 1.333 which further support the statistically significant result. Medium effect sizes of 0.6711 and 0.7789 were observed for the game score and response time respectively, which indicate a meaningful increase in the ASD participants' overall performance even though not all the categories were statistically significant. Note that for TD participants there were no statistically significant changes in all three categories even though the time to response had a medium effect size, 0.6702. Table 4-6 presents the pre-test and post-test performance measures.

Table 4-6: Overall Performance Measures Results

Participants		Pre	Post	T-test	
		Mean (SD)	**Mean (SD)**	***p-value***	**|*d*|**
ASD	**Highest score**	38.56 (16.82)	46.89 (5.06)	0.1313	**0.6711**
	Time to complete (seconds)	244.04 (74.74)	164.18 (39.93)	***0.0106**	***1.333**
	Response time (seconds)	3.44 (2.98)	1.72 (0.91)	0.0922	**0.7789**
TD	**Highest score**	47.56 (3.78)	48.67 (2.24)	0.2145	0.3579
	Time to complete (seconds)	192.90 (128.99)	169.67 (90.34)	0.32	0.2086
	Response time (seconds)	1.63 (0.76)	1.20 (0.48)	0.0608	**0.6702**

4.5.2 Game Score Measures Based on Gaze Prompt Speed

As mentioned in 4.3.5, the speed of the avatar's gaze prompt in the Bubble Popping game was increased by 2 ups each time the avatar provided a gaze prompt. Since the increment of the speed of gaze prompt in each turn was too small to be meaningfully analyzed individually, the avatar gaze prompt speed was clustered into five speed groups with a speed range of 10 ups in each cluster. For each group, the maximum score was 10 points. Figure 4-8 shows the performance in each speed group for both ASD and TD participants.

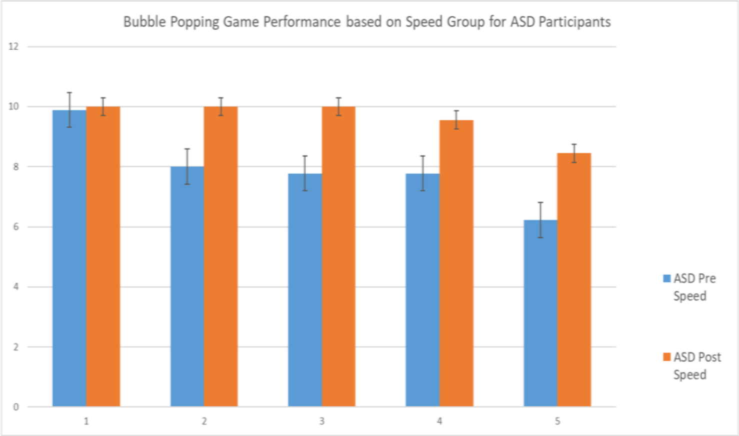


Figure 4-8: Performance comparison based on different speed grouping in pre and post-test for autistic participants in Bubble Popping Game.

Table 4-7 presents the results of statistical analysis using a t-test to compare the performance based on the different speed groups in the pre-test and post-test. The improvement in the performance was statistically significant for children with ASD (p = 0.0139). In the pre-test, the children with ASD were unable to keep up with the increase in speed of the avatar's gaze prompt as shown by their scores progressively declining from Speed Groups 1 to 5. However, in the post-test, the children with ASD achieved maximum possible scores in Speed Groups 1 to 3. For Speed Groups 4 and 5, their post-test performances were significantly better than their pre-test performances although they did not achieve the maximum possible scores. TD children continuously received maximum scores in Speed Groups 1-4 in both pre- and post-tests with minimal improvement in post-test for Speed Group 5. Again, consistent with the findings in the previous analysis of game performance, the result suggested that TD children were already performing at their highest level in all speed groups.

Table 4-7: Game Score Measures based on Speed Groups

Speed Group	**ASD**		**TD**	
	Pre	**Post**	Pre	**Pre**
Group 1	9.89 (0.33)	10.00 (0)	10.00 (0)	10.00 (0)
Group 2	8.00 (4.00)	10.00 (0)	10.00 (0)	10.00 (0)
Group 3	7.78 (4.41)	10.00 (0)	10.00 (0)	10.00 (0)
Group 4	7.78 (4.41)	9.56 (1.33)	10.00 (0)	10.00 (0)
Group 5	5.89 (4.48)	7.78 (3.56)	7.78 (3.67)	8.89 (1.96)
T-test	***p-value***	***0.0139**	***p-value***	0.3739
	|*d*|	***1.6050**	**|*d*|**	**0.5200**

4.5.3 Gaze Fixation

Gaze fixation was calculated from the defined ROI gaze points and gaze durations in MATLAB using one of the functions called "*fixation_detection.m*" available on EyeMMV toolkit [40]. The function used two spatial parameters and one temporal parameter. The first spatial parameter, *t1*, was used to initialize a fixation cluster. The second spatial parameter, *t2*, was used to establish consistency in the cluster by removing gaze points that were outside the threshold of the second spatial parameter. The temporal parameter defined the minimum duration for fixation. Any fixation cluster with a duration smaller than the defined value was not considered as fixation and was removed. The selection of these spatial and temporal parameters was based on the type of task that was carried out. In our analysis, we choose *t1* to be 1° of visual view and a minimum duration for fixation of 200 ms based on the study presented by Rayner [41] on reasonable human gaze fixation characteristic. As for *t2*, the threshold value was generated by the function by calculating the standard deviation from the fixation cluster.

To better understand the distribution of the participants' fixation on the avatar's face, we grouped the fixation points based on the ROI on the eye region and ROIs on other facial region. To get the fixation metrics for these ROIs, we ran the EyeMMV function for gaze points of each ROI separately. For example, to get the number of fixation points on avatar's eye region, we used gaze points corresponding only to the avatar's eye region, and to get the number of fixation points on other facial region of the avatar, we added the gaze points from the five passive ROIs; forehead, right ear, left ear, nose and mouth (as explained in 4.3.4 and in Figure 4-3). Table 4-8 represents the total fixation points on the avatar's face and normalized fixation on the avatar's eye region and other facial features.

Table 4-8: Results for gaze fixations on avatar's face

Participants		**Pre**	**Post**	**T-test**	
		Mean (SD)	**Mean (SD)**	***p-value***	**|*d*|**
ASD	**Total Face Fixation**	160.33 (46.29)	119.22 (46.95)	***0.0056**	***0.8914**
	Normalized Eye Fixation	0.42 (0.25)	0.60 (0.15)	[1]0.6546	[1]0.2688
	Normalized Other Facial Features Fixation	0.58 (0.25)	0.40 (0.15)	**[2]*0.0266**	**[2]*1.0474**
TD	**Total Face Fixation**	139.33 (104.66)	131.78 (74.76)	0.6700	0.0830
	Normalized Eye Fixation	0.63 (0.22)	0.59 (0.24)	[1]0.1876	[1]0.3556
	Normalized Other Facial Features Fixation	0.37 (0.22)	0.66 (0.24)	[2]0.8766	[2]0.0267

[1] p-value and Cohen's D value calculated using actual fixation points on avatar's eye region

[2] p-value and Cohen's D value calculated using actual fixation points on avatar's other facial features

The normalized result represents the ratio of the fixation points on the eye region to the fixation points on other facial features on the avatar's face. There was a statistically significant increase (p = 0.0056) in the total fixation points on the avatar’s face region for children with ASD. However, there was almost no change in the total fixation points on the avatar’s face for the TD children with low effect sizes that indicated trivial differences in the TD eye gaze fixation.

4.6 Discussion

We designed a novel VR gaze system, InViRS, to assess and teach skills related to two core features of joint attention: gaze sharing and gaze following in children with ASD. When designing the modules for InViRS, we wanted InViRS to accommodate the diverse learning abilities of autistic individuals since ASD is a spectrum disorder. Taking this into consideration, we designed and implemented the Game Adaptation Controller and the Assistive Avatar Module. The real-time use of eye gaze and game performance data in the Game Adaptation Controller created a personalized learning experience for children with ASD. Using the same real-time data in a supervisory logic embedded within the Avatar Assistive Module allowed InViRS to provide individualized hints or assistance when users were unable to progress in the tangram puzzle game.

We have successfully completed a pilot study using InViRS. In this study, children with ASD and TD children completed avatar-initiated RJA prompts in two games, one designed as a pre and post-test evaluation (Bubble Popping game) and one designed to allow real-time assistance and difficulty modification to prompt skill acquisition (Tangram Puzzle game). Gaze sharing was established by the avatar waiting for the participant to look its eye region before shifting its gaze toward the target. Gaze following was measured through the ability of the participant to correctly look at the object that was targeted by the avatar.

Based on the results and analysis presented above, we believe that this system has the potential to help children with ASD interpret important communicative gaze-based information as part of social interactions. Regarding gaze following, the overall performance of children with ASD improved as based on their higher game scores and shorter response times after practice with InViRS. This replicate other findings in the literature indicating that adaptive systems can enhance the learning experiences of people with ASD [42]. Regarding gaze sharing, children with ASD looked more frequently at the avatar’s eye region in the post-test as demonstrated by an increase in the ratio of fixation on the avatar’s ROI compared to other facial ROIs. This suggests that the assistive mechanism (LTM) embedded in the practice Tangram Puzzle games positively encourages the children with ASD to share their gaze with the avatar. This is consistent with the work [43, 44] supporting the use of a VR-system to assist individuals with ASD in shifting their attention to the desired object or event of interest. Results also suggest that the children with ASD learned that the avatar’s gaze communicated important non-verbal information with regard to the direction that they need to follow, as they spent less time looking for non-verbal prompts from other facial ROIs and more frequently directed their gaze at the avatar's eye ROI over time. However, even after gaze sharing was established, gaze following was still challenging, especially when the gaze prompt was quickly administered.

We also found important and persistent between-group differences based upon the speed with which gaze prompts were administered. Participants with ASD showed significant improvement in their performance in all speed groups. This statistically significant improvement indicated that InViRS was able to help children with ASD to adapt and respond to the changes in gaze prompts speed. However, relative to TD participants, it was harder for participants with ASD to correctly follow the avatar's gaze when it was quickly administered, even after they knew to look at the avatar's eye ROI. Looking at the pre-test results presented based on the different speed groups, participants with ASD scored relatively low in the higher speed group while TD participants showed consistently high performance across all speed groups. Furthermore, increasing the speed of the gaze prompts also encouraged the participants to respond to each gaze prompt faster. Faster response time to gaze prompts could indicate a more efficient joint attention ability. As previously reported in [44, 45], response time in a joint attention prompt were correlated with verbal intelligence [45] and ability to process social information [44]. It is also interesting to report that in the highest speed group, both ASD and TD participants did not receive full score, which could indicate that the avatar's gaze prompt speed in the highest speed group was hard to process.

The promising results of the current study further support InViRS as a system capable of tracking game data in varying configurations, accumulating game performance measures, adaptively changing the difficulty level while simultaneously interacting with participants and providing real-time feedback. As presented in the previous sections, we were able to see the differences in the performance measures and gaze data captured by InViRS, which characterize the discriminating gaze behaviors between autistic participants and TD participants. We compared the results between children with ASD and the TD children to establish any meaningful differences in the performance and gaze patterns. Our findings that the children with ASD exhibit atypical gaze patterns are consistent with other works on gaze related study of autistic individuals [3, 4, 44, 46]. For examples, in our study we found that children with ASD had lower ratio of fixation on eye compared to other facial features which was consistent with what was observed in [4], and they took longer time to respond to gaze prompts that was also found in [44, 46].

Although the results discussed above show promise, it is important to highlight the limitations of the study and important targets for future research. First, it was a short study with a relatively small sample size. A longitudinal study with a larger sample size would enable more complex analyses of InViRS's assistive capabilities and its impact. However, we believe that these preliminary results provide motivation and justification for a resource-intensive longitudinal study in the future. Next, there was no control group for this study. While it is not uncommon to not have a control group for a preliminary evaluation of a new system, we plan to include a control group in our future study to further assess the impact of InViRS in improving joint attention. Additionally, it will be interesting to explore the use of different facial expressions in RJA and its effect on children with ASD for joint attention tasks. It will also be beneficial to evaluate system functionality across different game types other than the two types of games we have used in this work. Finally, generalizability of the skills learnt in InViRS needs to be demonstrated in real-world situations. However, despite these limitations, results from the pilot study showed the potential of InViRS in improving both gaze sharing and gaze following skills in children with ASD. To our knowledge, this is the first such system and study that systematically manipulated these important components of joint attention skill. In addition, InViRS allowed measurement of several quantitative task-relevant metrics and provided real-time feedback to the participants to help them work on their RJA skills.

References

[1] MJ Maenner et al., "Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016." MMWR Surveill Summ 2020;69 (No. SS-4):1–12.

[2] JP Leigh, J Du, "Brief Report: Forecasting the Economic Burden of Autism in 2015 and 2025 in the United States." J Autism Dev Disord. 2015 Dec;45(12):4135-9.

[3] TW Frazier et al., "A Meta-Analysis of Gaze Differences to Social and Nonsocial Information Between Individuals With and Without Autism." Journal of the American Academy of Child and Adolescent Psychiatry. 2017; 56(7):546-555.

[4] E Thorup et al. "Altered gaze following during live interaction in infants at risk for autism: an eye tracking study." Molecular autism 7.1 (2016): 12.

[5] Bottema-Beutel K. "Associations between joint attention and language in autism spectrum disorder and typical development: A systematic review and meta-regression analysis." Autism Res. 2016;9(10):1021-1035.

[6] Mundy, P., & Newell, L. "Attention, Joint Attention, and Social Cognition. Current Directions in Psychological Science", 16(5), 269–274. 2007.

[7] Vaughn DL, Nye C. "Joint attention interventions for children with autism spectrum disorder: a systematic review and meta-analysis." Int J Lang Commun Disord. 2016;51(3):236-251.

[8] KA Murza et al., "A systematic review and meta-regression analysis of social functioning correlates in autism and typical development." Autism Res. 2019;12(2):152-175.

[9] Z. Zheng et al., "Design of an Autonomous Social Orienting Training System (ASOTS) for Young Children with Autism," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 668–678, 2017.

[10] S Ramdoss et al., “Computer-based interventions to improve social and emotional skills in individuals with autism spectrum disorders: A systematic review”, Developmental Neurorehabilitation, 15:2, 119-135 2012.

[11] S. Parsons, and S. Cobb, "State-of-the-art of virtual reality technologies for children on the autism spectrum." European Journal of Special Needs Education 26.3 (2011): 355-366.

[12] U. Lahiri, Z. Warren, and N. Sarkar, “Design of a gaze-sensitive virtual social interactive system for children with autism,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 4, pp. 443–452, 2011.

[13] U. Lahiri et al., “Design of a virtual reality based adaptive response technology for children with Autism,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 1, pp. 55–64, 2013.

[14] 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 Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 4, pp. 665–675, 2015.

[15] G. E. Little et al. “Gaze contingent joint attention with an avatar in children with and without ASD.” 2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob). IEEE, 2016.

[16] N Caruanaet al. “Joint attention difficulties in autistic adults: an interactive eye-tracking study.” Autism, 2017.

[17] M Courgeon et al. “Joint attention simulation using eye-tracking and virtual humans.” IEEE Transactions on Affective Computing 5.3 (2014): 238-250

[18] V. Yaneva et al., “Detecting High-Functioning Autism in Adults Using Eye Tracking and Machine Learning,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 6, pp. 1254–1261, 2020.

[19] Tanaka, J.W., Sung, A. “The “Eye Avoidance” Hypothesis of Autism Face Processing”. J Autism Dev Disord 46, 1538–1552 (2016).

[20] Kylliäinen, Anneli, et al. “Affective–motivational brain responses to direct gaze in children with autism spectrum disorder.” Journal of Child Psychology and Psychiatry 53.7 (2012): 790-797.

[21] Amat, A. Z. et al., “Design of an assistive avatar in improving eye gaze perception in children with ASD during virtual interaction.” In international conference on universal access in human-computer interaction (pp. 463-474). Springer, Cham, 2018, July

[22] Unity Website https://unity3d.com/unity

[23] Tobii EyeX https://gaming.tobii.com/products/

[24] A Gibaldi et al., “Evaluation of the Tobii EyeX Eye tracking Controller and Matlab toolkit for research. Behavior research methods”, 2017, 49(3), 923-946

[25] Tobii Unity SDK http://developer.tobii.com/tobii-unity-sdk/

[26] Maya, https://www.autodesk.com/education/free-software/maya

[27] M Tomasello et al., “Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis”, 2006.

[28] SV Shepherd, “Following gaze: gaze-following behavior as a window into social cognition.” Front. Integr. Neurosci. 4:5.

[29] Atsushi Senju and Gergely Csibra. “Gaze following in human infants depends on communicative signals.” Current Biology 18.9 (2008): 668-671.

[30] Hietanen, Jari K. “Does your gaze direction and head orientation shift my visual attention?.” Neuroreport 10.16 (1999): 3443-3447.

[31] S Johnson et al., “Whose gaze will infants follow? The elicitation of gaze-following in 12-month-olds.” Developmental Science 1.2 (1998): 233-238.

[32] M. E. Libby et al., “A comparison of most-to-least and least-to-most prompting on the acquisition of solitary play skills”, Behavior Anal. Pract., vol. 1, pp. 37-43, 2008.

[33] AS Polick et al., “A comparison of general and descriptive praise in teaching intraverbal behavior to children with autism.” Journal of Applied Behavior Analysis 45.3 (2012): 593-599.

[34] H Waddington, et al., “Three children with autism spectrum disorder learn to perform a three-step communication sequence using an iPad®-based speech-generating device”, International Journal of Developmental Neuroscience, Volume 39, 2014, Pages 59-67, ISSN 0736-5748.

[35] Finke, Erinn H., et al. “Effects of a least-to-most prompting procedure on multisymbol message production in children with autism spectrum disorder who use augmentative and alternative communication.” American journal of speech-language pathology 26.1 (2017): 81-98.

[36] Yanardağ, Mehmet, et al. “The effects of least to most prompting procedure on teaching basic tennis skills for children with autism.” Kinesiology (2011).

[37] Lord, Catherine, et al. “The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism.” Journal of autism and developmental disorders 30.3 (2000): 205-223.

[38] M Rutter et al., “Social Communication Questionnaire”, 2003. Los Angeles, CA: Western Psychological Services, 2003.

[39] Constantino, John N., and Christian P. Gruber. “Social responsiveness scale: SRS-2.” Torrance, CA: Western Psychological Services, 2012.

[40] V Krassanakis et al., “EyeMMV toolbox: An eye movement post-analysis tool based on a two-step spatial dispersion threshold for fixation identification.” (2014).

[41] K Rayner, “Eye movements in reading and information processing: 20 years of research.” Psychological bulletin 124.3 (1998): 372.

[42] Lahiri, Uttama, et al., “A physiologically informed virtual reality based social communication system for individuals with autism.” Journal of autism and developmental disorders 45.4 (2015): 919-931.

[43] Bian, Dayi, et al., “Online engagement detection and task adaptation in a virtual reality-based driving simulator for autism intervention.” International Conference on Universal Access in Human-Computer Interaction. Springer, Cham, 2016.

[44] Falck-Ytter, Terje, et al., “Gaze performance in children with autism spectrum disorder when observing communicative actions.” Journal of autism and developmental disorders 42.10 (2012): 2236-2245.

[45] Van Hecke et al., “Infant responding to joint attention, executive processes, and self-regulation in preschool children.” Infant Behavior and Development 35.2 (2012): 303-311.

[46] L Chukoskie et al., “A novel approach to training attention and gaze in ASD: A feasibility and efficacy pilot study.” Developmental neurobiology 78.5 (2018): 546-554.

[47] IP Howard and Brian J. Rogers. “Binocular vision and stereopsis.” Oxford University Press, USA, 1995.

[48] Kenny L, et al., “Which terms should be used to describe autism? Perspectives from the UK autism community.” Autism. 2016 May;20(4):442-62.

[49] Battocchi, A. et al., “Collaborative puzzle game-an interface for studying collaboration and social interaction for children who are typically developed or who have Autistic Spectrum Disorder”, 7th ICDVRAT with ArtAbilitation. ICDVRAT/University of Reading, UK, 2008, 127-134.

[50] Zhang, L. et al., “Design and evaluation of a collaborative virtual environment (CoMove) for autism spectrum disorder intervention.”, ACM Transactions on Accessible Computing (TACCESS), 11(2), 2018, pp.1-22.

[51] Zhao, H. et al, “Design of a tablet game to assess the hand movement in children with autism.” In International Conference on Universal Access in Human-Computer Interaction, July 2017, (pp. 555-564), Springer, Cham.

This chapter was previously published in © 2021 IEEE. Reprinted with permission from A. Z. Amat, H. Zhao, A. Swanson, A. S. Weitlauf, Z. Warren and N. Sarkar, "Design of an Interactive Virtual Reality System, InViRS, for Joint Attention Practice in Autistic Children," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 29, pp. 1866-1876, 2021, doi: 10.1109/TNSRE.2021.3108351.

CHAPTER 5: DESIGN OF A VIRTUAL REALITY-BASED COLLABORATIVE ACTIVITIES SIMULATOR (VIRCAS) TO SUPPORT TEAMWORK IN WORKPLACE SETTINGS FOR AUTISTIC ADULTS

5.1 Abstract

Autistic adults possess many skills sought by employers, but may be at a disadvantage in the workplace if social-communication differences negatively impact teamwork. We present a novel collaborative virtual reality (VR)-based activities simulator, called ViRCAS, that allows autistic and neurotypical adults to work together in a shared virtual space, offering the chance to practice teamwork and assess progress. ViRCAS has three main contributions: 1) a new collaborative teamwork skills practice platform; 2) a stakeholder-driven collaborative task set with embedded collaboration strategies; and 3) a framework for multimodal data analysis to assess skills. Our feasibility study with 12 participant pairs showed preliminary acceptance of ViRCAS, a positive impact of the collaborative tasks on supported teamwork skills practice for autistic and neurotypical individuals, and promising potential to quantitatively assess collaboration through multimodal data analysis. The current work paves the way for longitudinal studies that will assess whether collaborative teamwork skill practice that ViRCAS provides also contributes towards improved task performance.

5.2 Introduction

Autism spectrum disorder (ASD) impacts social communication and interaction as well as behavior and sensory processing [1]. One in 44 children and 1 in 45 adults are diagnosed with ASD in the US [2] with more than 70,000 autistic children reaching adulthood each year [3]. Differences in social communication and interaction can disadvantage autistic adults as they attempt to secure and retain employment [4]. The unemployment rate for autistic adults is between 50% and 85%, higher than other types of disabilities [5]. This contributes to high lifetime care costs [6] partly due to unemployment [7]. Although preference varies between the person-first language (i.e., adults with autism) and the identity-first language (i.e., autistic adults) and we are respectful for both uses, we use identity-first language in this manuscript due to 1) a recent survey that reported a majority preference for identity-first language across the globe [8]; and 2) our stakeholder partners indicated that they wished to be referred as "autistic adults".

Autistic adults may have many workplace-relevant talents [9], such as attention to detail [10]. However, differences in communication and social interaction skills relative to colleagues without ASD (“neurotypical”) can impact employment opportunities that require a high level of teamwork [11]. In general, teamwork skills are associated with improved productivity and workplace performance [12], are among the core skills sought by employers, and can influence hiring decisions [13]. Companies such as Microsoft and Specialsterne have started using a non-traditional interview process to assess teamwork skills of autistic job candidates using Lego Mindstorm group projects [14] and Minecraft [15]. Therefore, supporting autistic adults to acquire work-relevant teamwork skills may contribute to not only job acquisition, but also improved workplace social communication skills [16], problem-solving skills [17], and self-confidence [18].

One way to assess and support teamwork skills development is simulation-based training (SBT), which enables individuals to engage in shared social, cognitive, and behavioral processes pertaining to a collaborative task [19]. Although existing SBT programs have positively impacted teamwork skills development [18], these programs can be tedious, resource-straining, and costly [20], thus driving the need for a technology-based solution. Over the last decade, the use of human-computer interaction (HCI) technology has shown promise by providing lower-cost, engaging interactions that can expand accessibility [21]. For example, Virtual Reality (VR) has been used to simulate real-world scenarios for skills training at a lower cost [22]. VR-based systems have shown promise for teaching both autistic and neurotypical individuals new social and technical skills [23]–[25]. However, conventional VR-based systems are unable to support the complex back and forth human-human interactions important for teamwork skills training. Additionally, overreliance on virtual interactions may limit generalizability and success in real world tasks [26].

Effective teamwork requires collaboration among individuals working together, making collaboration an important indicator of teamwork performance [27]. A collaborative virtual environment (CVE) extends the benefits of conventional VR technology by supporting multi-user interaction within the shared virtual space, allowing users to naturally communicate with each other [28], potentially increasing generalizability of learned skills to the real world. Several recent CVE-based interactions have been promising. CVE-based studies that focused on social communication skills reported improvements in emotion recognition skills and conversational skills in autistic users when they participate in a series of social scenarios with other users in CVEs [29], [30]. In another study, researchers compared social communication performance between CVE and face-to-face interactions and found that users in the CVE had more verbal exchanges compared to users in the physical group for the same task [31]. Previous CVE-based studies also explored non-verbal aspects of social communication such as joint attention skills [32], joint action skills [33], and imitation skills [34] where improvements in these skills were observed after completing multiple training sessions in laboratory settings. More recently, researchers found promising results by combining social communication skills and motor skills training for autistic children where they reported increased performance in both domains [28]. However, to our knowledge, there are no CVE-based studies that target social interactions within the employment landscape, which can differ from everyday social interactions.

Complex social skills such as teamwork can be challenging to assess [35]. Existing methods of assessment still rely heavily on human observations [29], [36]. Fortunately, studies on collaborative learning and communications can be leveraged to objectively assess teamwork skills [37]. Furthermore, advancements in HCI and sensors technologies have paved the way for the use of multimodal data to provide a reliable assessment of human behavior [38], through quantitative measures of several dimensions of collaboration [39].

Motivated by the need to support autistic adults to succeed in the workplace as well as the potential of CVE as a platform to train teamwork skills, we present in this paper the design, development, and initial feasibility results of a novel Virtual Reality-based Collaborative Activities Simulator (ViRCAS). ViRCAS is a virtual simulator that allow two individuals (one autistic adult, one neurotypical [NT]) in physically distributed locations to participate in interactive activities over the network, with the goal of fostering and measuring change in teamwork skills. ViRCAS is designed as a desktop-based CVE, the least immersive form of VR [40], to prevent cyber sickness that could be caused by immersive VR such as nausea, and dizziness [41]. The primary contributions of this work are: 1) a new CVE- based teamwork skills practice platform for two individuals; 2) a set of stakeholder-driven collaborative tasks with embedded collaboration strategies; and 3) a framework for multimodal data analysis to assess collaboration.

The current work substantially expands our previous conference paper [42] in terms of 1) expansion of system interactivity: we incorporated audio and visual communication channels within the CVE that allow the users to see and talk to each other; 2) introduction of a new collaborative task: Task 3 in Section II-A-3 and the addition of difficulty levels in all tasks; and 3) classifying and assessing collaboration using multimodal data from a human-subject study. The remainder of the paper is organized as follows: Section II presents the system design and the system architecture. Section III describes the experimental setup followed by Section IV, which presents the results of the study. Finally, Section V discusses the results and addresses the potential and limitations of the current study.

5.3 Collaborative Virtual Environment (CVE) System Design

5.3.1 Collaborative Tasks Design Principles

5.3.1.1 Stakeholder-driven Universal Design of Collaborative Tasks:

We employed a participatory design process where we engaged with stakeholders and end-users from various backgrounds to design meaningful collaborative tasks: industry representatives from 2 companies, a certified behavioral interventionist, 2 career counselors from 2 vocational rehabilitation centers, and 3 autistic adults. Stakeholders were involved in both the design and development stages of the collaborative tasks. In the design stage, we conducted multiple discussion sessions with the stakeholders to select tasks that are collaborative and include interactions that are suitable in a workplace environment. For example, a puzzle game task can be collaborative, but might not involve workplace-related interactions. The collaborative tasks selection was driven by employment-related studies for autistic individuals: a) a PC

Assembly task [43], b) a Fulfillment Center task [44], and c) a Furniture Assembly task [45]. These tasks elicited teamwork-relevant behaviors between two users, could be designed at varying difficulty levels, and involved workplace-related interactions. Additionally, we incorporated universal design principles into our collaborative tasks to create a system that can be used by individuals with different abilities [46].

In the development stage, we recruited 3 autistic adults and 3 neurotypical adults. Each ASD-NT pair tested the initial version of the collaborative tasks while being observed by two expert behavioral interventionists with prior experience in real-world teamwork tasks, who then commented on task suitability and made suggestions for improvement to align with real-world supports. We made several changes based on feedback from interventionists and participants. First, they identified a need for a specific and structured task guide. As a result, we developed a tutorial level that provided step-by-step instruction in task components.

Second, participants found that the virtual objects were difficult to manipulate. To address this concern, we simplified the object manipulation function in three ways. First, we automated object rotations in both the PC Assembly and Furniture Assembly tasks. Second, we highlighted the target area or objects to guide the participants in all three tasks. Third, we removed the gear shifter button from the gamepad for a smoother driving experience in the Fulfillment Center task. Finally, participants said they sometimes were not sure what they needed to do. We therefore added visual cues that made it easier for participants to know where to go or which objects to move. All of these stakeholder-informed changes were applied prior to the next phase of work.

5.3.1.2 Collaboration dimensions for Collaborative Activities:

Based on literature related to dyadic interactions and collaboration [37], [39], we used the 9 dimensions of collaboration given in [39] into our tasks. These dimensions use both verbal and non-verbal communications that can be quantitatively measured to represent the quality of collaboration between the participants. Table 5-1 lists the dimensions and their definitions.

Table 5-1: Dimensions of Collaboration

No.	Dimensions	Definition (The task should allow...)
1	Sustaining Mutual Under- standing	Participants to share ideas and show mutual understanding.
2	Dialogue management (Turn-taking)	Participants to engage in back-and-forth communication and activities.
3	Information Pooling	Participants to share information with each other.
4	Reaching Consensus (Decision making)	Participants to discuss and agree with each other
5	Task Division	Participants to discuss and coordinate their actions within the task
6	Time Management	Participants to monitor and be aware of time restrictions in the task

7	Technical Coordination	Participants to handle technical dependencies of the task
8	Reciprocal Interaction	Participants to progress at the same pace
9	Individual Task Orientation	Participants to perform individual actions independently

5.3.1.3 Task Descriptions

First, we will describe the overview and setup of the collaborative activities simulator. Two participants in different physical locations accessed a shared virtual environment from their respective computers as illustrated in Figure 5-1. Each participant used the input device to interact with their virtual environment, a headphone with a microphone to communicate with their partner, and a webcam to see each other. After they were connected to the same virtual environment, participants could communicate with each other through an audio and video streaming component embedded within the virtual environment. They were asked to complete a tutorial and 2 difficulty levels of each task and no systematic differences were created between the roles, activities or equipment for Player 1 and Player 2 (see Table 5-2 for task and level descriptions).

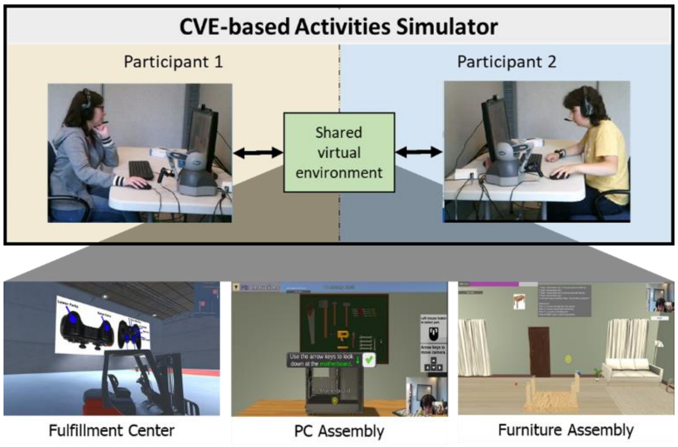
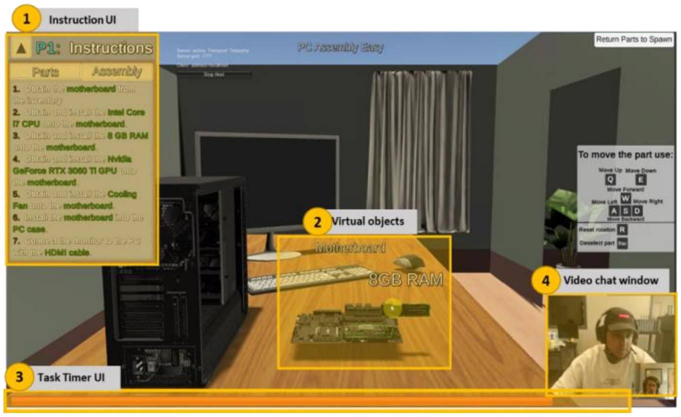


Figure 5-1: ViRCAS setup and snapshots of the three collaborative tasks.

For the PC Assembly task, both users were assigned the same role of putting together different computer hardware to build a computer. Users had different points of view of the working area as if they were located at different ends of the table. Once the participants completed the tutorial, in the Easy level, participants were given the assembly instructions, but each participant was given a different list of computer hardware. In the Hard level, participants were given mismatched assembly instructions with missing information and additional computer hardware to assemble. In both levels, participants had to exchange installation instructions and work together to place the hardware in the correct locations.

For the Fulfillment Center task, both participants needed to drive a forklift to pick up and deliver crates from a storage shelf to a collection area in a warehouse. Each forklift had different height capacity; one forklift could only raise its fork to medium height while the other forklift could lift the fork to a higher height. Participants were given a map that showed them where the crates were located. After the tutorial, both participants were given different lists of crates that they needed to pick up. In the next level, additional crates were placed at different heights.

In the Furniture Assembly task, participants had to work with each other to assemble various furniture pieces. After the tutorial, in the Easy and Hard levels, both participants needed to work together to assemble a coffee table and a bookcase, respectively. The variation in the type of furniture influenced the difficulty level of the task. In addition, participants were given assembly instruction in the Easy level, while only an image of a completed furniture in the Hard level.



1. Obtain the motherboard from the inventory
2. Obtain and install the Intel Core I7 CPU onto the motherboard
3. Obtain and install the 8 GB RAM onto the motherboard
4. Obtain and install the Nvidia GeForce RTX 3060 Ti GPU onto the motherboard
5. Obtain and install the Cooling Fan onto the motherboard
6. Install the motherboard into the PC case
7. Connect the monitor to the PC with the HDMI cable

Figure 5-2: Example of Regions of Interest (ROIs) for PC Assembly Task.

Table 5-2: Collaborative Task Levels

Tasks	Tutorial	Easy Level	Hard Level
PC Assembly	• Step-by-step instruction to familiarize participants with computer parts and controller	• Same instruction manual • Different components in their inventory • 7 steps to complete installation of the PC	• Instructions for each player contain missing information • Additional components in their inventory 12 steps to complete installation of the PC
Fulfillment Center	• Participants take turns driving the forklift When one participant is driving the forklift, the other participant will provide verbal instruction on where to pick up the crate	• Participants were given their own forklift to drive • Participants pick up one crate each and drop it off at the designated location	• Participants need to pick up 3 crates each • Crates assignment mismatch the forklift height capacity
Furniture Assembly	• The same instruction is given to both participants • Move 4 objects in the living room to a dedicated location	• Instructions with different information given to each participant • 5 furniture parts to assemble	• No instruction given, only a picture of the completed furniture • 9 furniture parts to assemble

5.3.2 ViRCAS Architecture

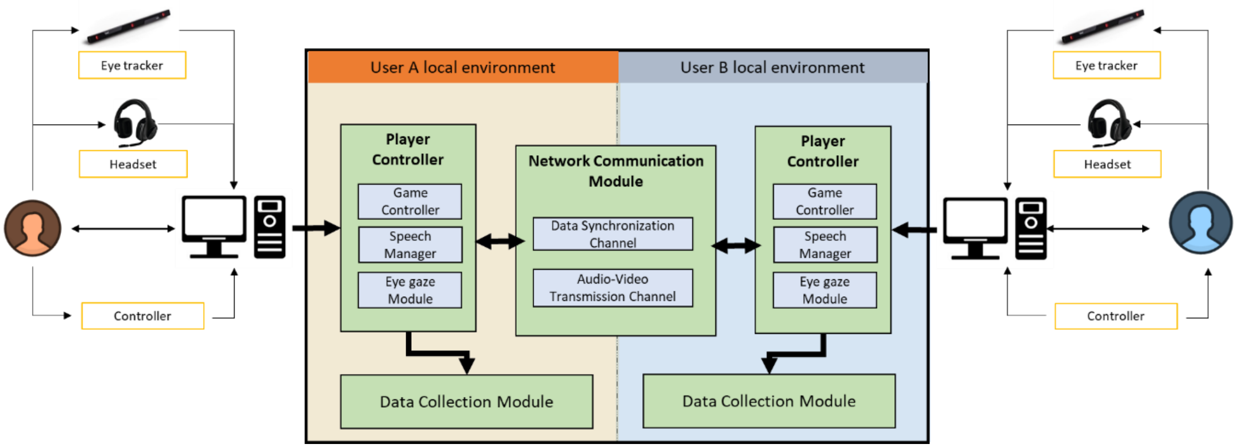


Figure 5-3: Architecture of the collaborative system.

1) *Input Devices:* One principle of universal design is perceptible information [46], which supports multiple methods of communication between users and the system. We employed three types of input devices with varying characteristics, with one device for each of the tasks as presented in Table 5-3 to explore their benefits. In the PC Assembly task, the participants used a keyboard and mouse to move the virtual hardware. In the Fulfillment Center task, the participants used a Logitech Gamepad [47]. Participants used the directional pad to drive the forklift in the virtual warehouse and the directional button to change the height of the fork when picking up a crate. In the Furniture Assembly task, the participants used a haptic device [48] for greater immersion.

Table 5-3: Input Device Specifications

Specification	Keyboard and Mouse	Gamepad	Haptic Device
Ease of use	Simple to use	Require minimal practice	Require more practice
Realism/ Immersion	No feedback to users	No feedback to users	Users can ’touch’ and feel the ’weight’ of the virtual object
Cost	Low-cost	Low-cost	High-cost
Task	PC Assembly	Fulfillment Center	Furniture Assembly

2) *CVE Modules and Communication Network:* Figure 5-3 illustrates the system interaction diagram and architecture. The ViRCAS was created using a multi-platform game development software, Unity [49]. The Network Communication Module handles the connection of two participants to the same virtual environment. This module also manages real-time audio and video interaction. Virtual objects’ synchronization was achieved using a Unity plugin called Mirror [50], while the audio and video data streaming were accomplished using WebRTC [51]. Task-related data are transmitted between the two

computers in packets. Mirror uses Transmission Control Protocol (TCP) to send information between the two computers. TCP ensures data transmitted from the source are correctly delivered to the target, in the right order, resulting in a synchronized shared environment. Although TCP assures data delivery to the target, the latency is slightly higher and could result in delay. However, for ViRCAS, the latency does not significantly affect task interaction since our tasks do not require instantaneous updates. WebRTC uses User Datagram Protocol (UDP) for audio and video transmission which prioritizes latency over data accuracy.

Next, the Player Controller component manages the use of multiple peripheral devices by participants to interact with the virtual environment. Task-related data collected in this module are sent over to the Network Communication Module.

There are three sub-modules within the Player Controller component. First, the Game Controller manages input device manipulation of virtual objects. The Player Controller component manages the input devices and keeps track of the task time and task progression. Second, the Speech Manager processes participants' speech. The spoken words from both ASD and NT participants are transcribed into text in real-time using Microsoft Azure's Speech-to-Text service [52]. Although we did not conduct our own evaluation of Azure's performance of speech-to-text, Alibegovic et al. reported that Azure's Speech-to-Text API had the lowest word error rate (WER) compared to other speech transcription APIs for general English speech transcriptions for individuals with accents [53]. Azure has also been used in multiple studies with individuals with disabilities including autism with good transcription performance [54], [55]. In Unity, we created a continuous listener function that captures any speech and sends it over to Azure API. Upon receiving the data, Azure proceeds to transcribe each word it received and grouped the words as one utterance. One utterance ends when silence was detected or a maximum of 15 seconds of audio was processed [52]. We can determine the number of words used in each utterance and the duration of the utterance with the transcribed speech.

The third and final sub-module is the Eye Gaze Module that detects participants' eye gaze on the computer screen using a TobiiEyeX eye tracker and the Tobii Eye Tracking Windows application for calibration [56]. Gibaldi et al. reported that the overall performance of TobiiEyeX in terms of the native calibration performance, accuracy, latency, and sampling frequency was acceptable for active fixation evaluation in 3D environment like ViRCAS [57]. We utilized a Tobii Unity Eye Tracking SDK [56] to: 1) continuously capture gaze points, and 2) capture gaze fixations on pre-defined region of interests (ROIs) and virtual objects when a gaze duration of approximately 200 ms is detected. Figure 5-2 presents an example of ROIs for the PC Assembly task. We calibrated participant gaze before each experimental session to improve the accuracy of the gaze points. Finally, the controller data, speech data, detected gaze points, and the ROIs were recorded together with the timestamps and sent to the Data Collection Model.

In the Data Collection Module, we captured and recorded multimodal data from each participant in a one-second interval. However, if multiple utterances were detected in one second, the transcribed speech would be logged in multiple sequences with the same timestamp. As for eye gaze data, since the eye tracker captures up to 60 gaze points in 1 second, we calculated and recorded the average point in the log file. Data from both participants were consolidated into a single log file for easy analysis.

3) *Multimodal Data Mapping:* An important contribution of ViRCAS is its capacity to capture multimodal data from both participants as quantitative measures of teamwork and collaboration. First, we captured participants' speech using dedicated microphones that were connected to the computer of each participant. From speech data, we derived the (1) transcribed speech, and (2) the number of words per sentence to capture verbal communication. Second, eye gaze data provided important information on non-verbal communication in collaborative activities. For example, when a participant mentioned an object's name, the other participant could re- spond by looking at the object or read information from the instruction. The gaze data gave us the (3) location of the gaze in xy-coordinate on the screen, and the (4) ROIs which could be either virtual objects or an area on the screen they are looking at. Third, we captured the input device data to detect collaborative activities, which were (5) input device manipulation such as button clicks or position of the haptic device, (6) name of the virtual objects, and (7) movements of the objects. All the data were collected together with the (8) timestamp, and (9) player label (either Player 1 or Player 2). These data were used to identify the collaboration dimensions that were defined in Section 5.3.1.2.

4) *Task Management using Finite State Machine:* We de- signed a finite state machine (FSM) applicable to all three tasks in the Player Controller module to manage seamless state transitions for two participants as they navigate through the task. Figure 5-4 presents the FSM used for all three collaborative tasks.

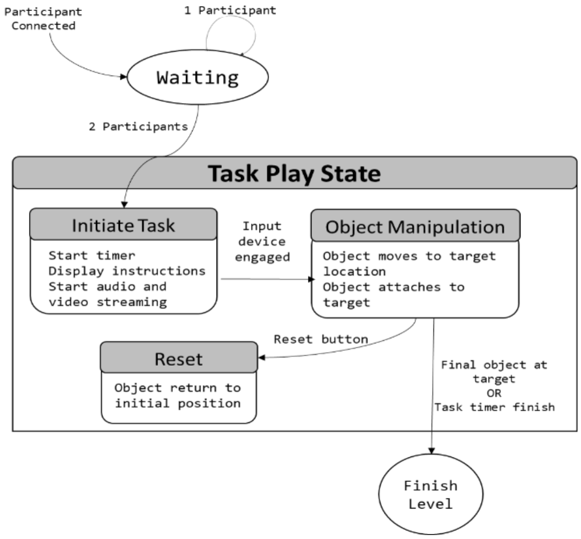


Figure 5-4: Generic finite state machines for all tasks.

5.4 Experimental Design

We conducted a feasibility study to 1) assess the usability and acceptability of ViRCAS for autistic and NT individuals; 2) assess the ability of the tasks to support teamwork; 3) measure various dimensions of collaboration during interaction; and 4) compare collaboration patterns of both autistic and NT individuals. The experiment was conducted with two groups of paired participants; one group of ASD and NT pairs (ASD-NT group) and one group of NT and NT pairs (NT-NT group). This study was approved by the Institutional Review Board at Vanderbilt University (IRB number: 161803).

The experiment was run on two standard desktop computers, with the same specifications, equipped with Windows 10 Education, with an Intel Xeon E5-1650 CPU @3.20GHz, and 16GB of RAM, with a 28 inch LCD with 1920 × 1080 resolution running at 60 Hz.

5.4.1 Participants and Protocol

We recruited 6 autistic individuals and 18 NT individuals (ages: 16 – 30 years; mean age: 23.4 years) to participate. Participants with ASD were recruited from a large research registry maintained by the Vanderbilt Kennedy Center of individuals previously diagnosed with ASD by licensed clinical psychologists. The NT participants were recruited from the local community through regional advertisement. We then divided the participants into two groups: 6 ASD-NT pairs, and 6 NT-NT pairs. Table 5-4 shows the characteristics and current level of ASD symptoms of all participants as measured by the Social Responsiveness Scale, Second Edition (SRS-2) [58]. Note that SRS-2 T-scores of 66 and above reflect at least moderately elevated symptoms of ASD, while T-scores of 59 and below reflect little-to-no evidence of ASD.

Table 5-4: Characteristics of Participants

Participants	ASD (N = 6)	NT (N = 18)
	Mean (SD)	**Mean (SD)**
Age	22.55 (1.8)	24.25 (2.1)
Gender (% male)	55.6%	55.6%
Race (% White, % African American)	80%, 16.7%	83%, 5.6%
Ethnicity (% Hispanic)	33.3%	11.1%
SRS-2 T-score	75.22 (7.38)	45.64 (16.12)

SRS-2: Social Responsiveness Scale, Second Edition

Each pair of participants attended a one-time visit to the laboratory that lasted approximately 90 minutes. They were seated in two different experiment rooms and accessed ViRCAS from local area network (LAN) that ensured data security and privacy. Before the participants began their session, consents and assents from the participants' guardians and the participants themselves were obtained, respectively. Participants completed all levels (i.e., Tutorial, Easy, and Hard) of the PC Assembly, Furniture Assembly, and Fulfillment Center task.

5.5 Results

5.5.1 Acceptability of the Collaborative Tasks and Input Devices

We administered 24 written questions rated with a 10 point-Likert scale (1 – very uncomfortable, 10- very comfortable) to get participant feedback on the collaborative tasks, input devices, and system acceptability (see Table 5-5). Results indicated that all tasks were acceptable to both groups with average score values of 7 or higher. Autistic participants preferred the gamepad the most, while NT participants preferred keyboard and mouse. The haptic device was the least preferred device in both groups. However, participants did express positive verbal comments on the "touch" sensation they felt while using the haptic device.

Table 5-5: Questionnaire Score

Questions	ASD (N = 6) Mean (SD)	NT (N = 18) Mean (SD)	t-stat	p-value
Collaborative Tasks				
How confident did you feel throughout the task?	7.94 (2.50)	7.00 (2.72)	1.179	0.2723
How comfortable did you feel overall with the task?	8.38 (1.78)	7.48 (2.22)	1.349	0.2070
How comfortable did you feel interacting with your partner?	9.52 (1.23)	9.34 (1.16)	0.651	0.5267
How comfortable did you feel when the task was challenging?	8.47 (1.88)	6.95 (2.43)	**4.306**	**0.0007***
Input Devices				
How comfortable did you feel using the haptic device to move parts around?	6.66 (3.07)	4.88 (2.51)	1.280	0.2414
How comfortable did you feel using the keyboard and mouse to move parts around?	7.16 (2.31)	8.11 (1.81)	-0.910	0.393
How comfortable did you feel using the gamepad to move parts around?	8.83 (1.60)	7.44 (2.57)	1.558	0.1417
ViRCAS System				

How much do you agree with the following: “Practicing with this system would help me work with others better.”	9.50 (1.22)	8.11 (2.39)	1.840	0.0822
How much do you agree with the following: “If it was available, I would use this system to practice my teamwork skills”	9.66 (0.81)	7.00 (3.04)	**3.367**	**0.0028***

* p-value < 0.05

A *t*-test for unequal sample size revealed two questions that were statistically significant as indicated in Table V. The first question was related to comfort level when the task was challenging. Autistic participants gave a rating of 8.47, while NT participants rated it at 6.95 (p-value < 0.001). The second question was related to whether participants would use the system to practice teamwork, if available. We found that autistic participants rated themselves as more likely to use the system than NT participants (p-value < 0.001).

5.4.1 Dialogue Acts Classification

We analyzed the transcribed speech data to better understand the context of the conversation. Table 5-6 lists the annotation scheme adapted from a verbal behavior coding scheme used to classify the speech [59]. For this study we used function words instead of individual word to generalize the labeling to all three tasks [60]. For example, in the sentence, “So it wants you to put that into the cpu”, instead of focusing on the word “cpu”, we looked at the words “wants”, “put”, and “that” to classify this sentence as “Inform”. Two annotators labeled the transcribed conversation between the participants using the coding scheme and reached an inter-annotator agreement of 95%. The annotators reconciled their differences to reach a final agreement of 100%.

Table 5-6: Dialogue Acts Definitions

Label	Definition	Example
Acks	Indicate agreement or acknowledge	‘I know’, ‘you’re right’, ‘okay’, ‘yeah’, ‘yup’, ‘cool’, ‘uh-huh’, etc.
Desc	Describe action or intention, decision making	Personal statements of opinion or non-opinion. ‘I think’, ‘I feel’, ‘I believe’, ‘I mean’, etc.
Neg	Disagree, confused, negative statements	‘No I don’t need this one’, ‘I don’t think this is the right one’, ‘no’, ‘um, I’m not sure’, ‘I don’t think so’, ‘oh no’
Pos	Positive feedback from one participant to another.	‘Well done’, ‘good job’
Ques	Questions	‘What do you see?’, ‘can you try W?’
Read	Any indication that the participant is reading task instructions.	Mine says to select the 8 gigabyte RAM'
Inform	Inform, instruct. Action directive statements or statements of instruction from one participant to another.	‘Try moving it more to the right’, ‘and then backwards’, ‘let’s see’, ‘mine has me moving’, ‘let me try’
Conv	Conventional pleasantries	‘thanks’, ‘thank you’, ‘sorry’, ‘my bad’
Out	Uninterpretable. When utterance is incomplete or does not make sense to the coder	‘the. ’, ‘it said an end then snow ’

We used an unpaired t-test to evaluate any differences in the labeled utterances between the two groups. As shown in Table 5-7, we found statistically significant differences in five types of dialogue acts between the ASD- NT group and NT-NT group. Pairs in the NT-NT group uttered more acknowledgements (“Acks”, p-value < 0.001), used more negative words (“Neg”, p-value < 0.001), asked more questions (“Ques”, p-value < 0.001), and instructive utterances (“Inform”, p-value < 0.001), while pairs in ASD-NT group used descriptive words more (“Desc”, p-value < 0.001) compared to NT-NT group. Figure 5-5 illustrates noticeable differences in the dialogue acts percentage between the groups.

Table 5-7: Dialogue Acts Classification Frequency

Labels	**ASD-NT** **Mean (SD)**	**NT-NT** **Mean (SD)**	**t-stat** **(p-value)**
Acks	12.142 (82.304)	15.143 (58.924)	**1.889 (0.031)***
Desc	15.107 (153.188)	10.375 (67.002)	**-2.387 (0.009)***
Neg	4.214 (15.553)	8.393 (34.897)	**4.402(1.389e-5)***
Pos	2.321 (6.986)	2.518 (6.509)	0 .4001 (0.345)
Ques	3.625 (16.420)	6.107 (17.406)	**3.194 (0.0009)***
Read	1.857 (7.761)	2.589 (7.083)	1.422 (0.079)
Inform	5.518 (58.036)	12.286 (120.68)	**3.788 (0.0001)***
Conv	0.714 (0.826)	0.964 (1.089)	1.351 (0.09)
Out	1.964 (3.344)	1.582 (2.989)	-1.132 (0.13)

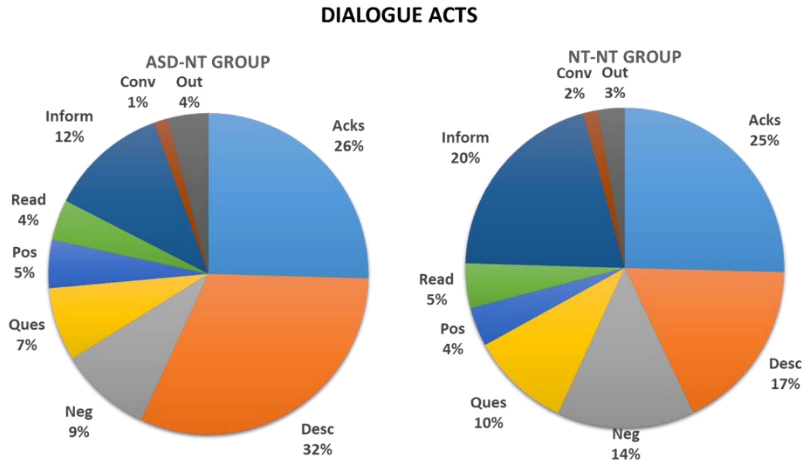


Figure 5-5: Dialogue acts percentages in ASD-NT group and NT-NT group.

5.4.2 Utterances Analysis

We analyzed the utterances by grouping the number of utterances into Easy and Hard levels as we wanted to observe the impact of increasing the difficulty level on collaboration. We found that the number of utterances increased as the difficulty level increased for all participants as shown in Figure 5-6. The results were divided into four groups; ASD, NT, NT 1 (NT-NT group), and NT 2 (NT-NT group), to distribute the number of participants per group evenly, i.e., N = 6 as we wanted to compare number of utterances of autistic individuals to NT participants. A t-test showed statistically significant increase in utterances for participants in the NT-NT group across all tasks (p-value < 0.001), and a statistically significant increase for the ASD-NT group only for the Furniture Assembly Task (p-value < 0.001).

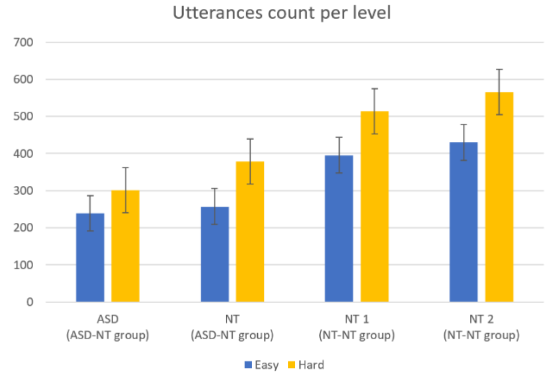


FIGURE 5-6: NUMBER OF UTTERANCES INCREASED FOR ALL PARTICIPANTS BETWEEN EASY AND HARD LEVEL.

5.4.3 Gaze Duration Results

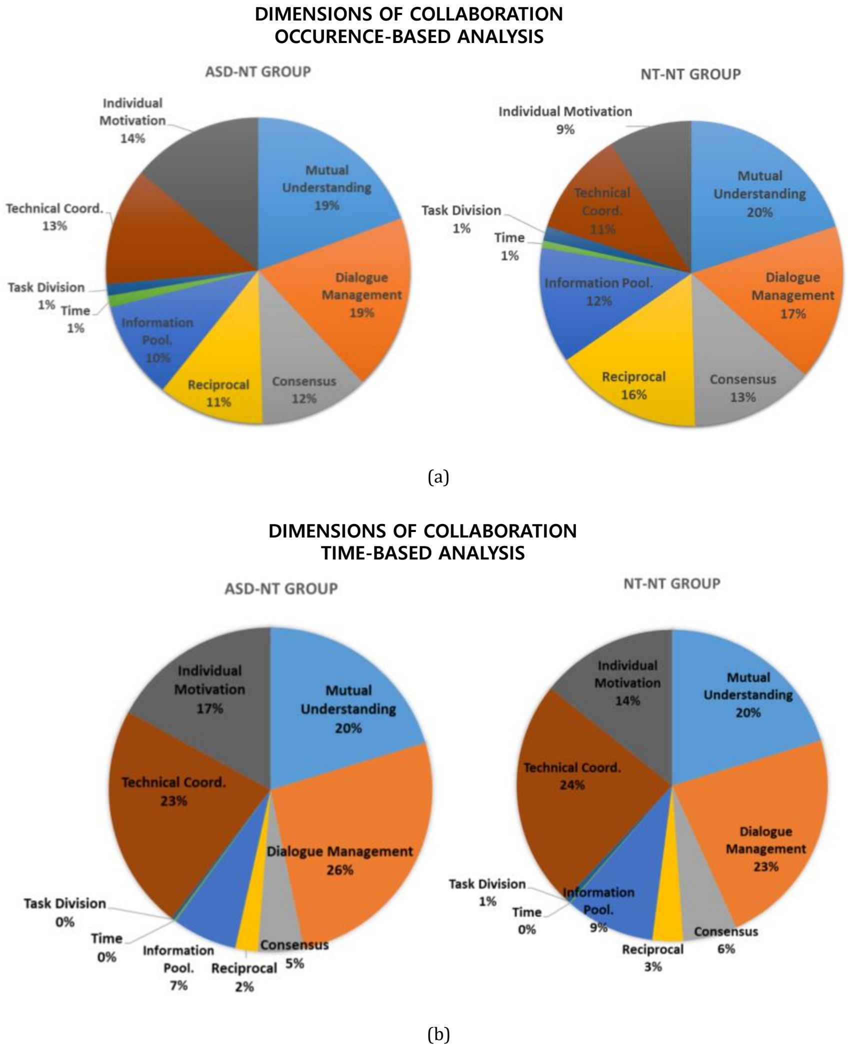
We analyzed the participants' gaze by calculating the du- ration of the gaze on the ROIs. Gaze duration that lasted approximately 250 ms was considered a "gaze fixations." Table 5-8 compares the average gaze fixations duration between autistic and NT participants. An unpaired t-test showed statistically significant differences in gaze fixations duration for the Fulfillment Center task; NT participants gazed 3 times longer at the virtual objects compared to participants with ASD (p-value < 0.05). Other tasks did not show any significant differences.

Table 5-8: Participants' Gaze Duration

Tasks	**ASD** **Mean (seconds)**	**NT** **Mean (seconds)**	**t-test** **t-stats** **(p-value)**
PC Assembly	74.506	68.501	0.144 (0.445)
Fulfillment Center	47.895	131.685	**2.756 (0.012)***
Furniture Assembly	67.556	82.879	0.462 (0.328)

5.4.4 Observation of Dimension of Collaboration

To determine whether ViRCAS captured the dimensions of collaboration (Table 5-1) from the multimodal data, we first labeled the occurrence of each dimension from every 30 s of sampled data and then computed the occurrence and the duration of each dimension within the sampled 30 seconds. The percentage of the dimension occurrence are presented in Figure 5-7(a), and duration percentage in Figure 5-7(b). Both ASD-NT pairs and NT-NT pairs showed similar collaborative patterns in both analyses. However, pairs in the NT-NT group showed more reciprocal interaction and less technical coordination and individual motivation compared to the pairs in the ASD-NT group. In both groups, the occurrences of task division and time management dimensions were very low compared to the other dimensions at only 1% occurrence. In the occurrence-based analysis, dialogue management and mutual understanding occurred most frequently, while in the time-based analysis, participants in both groups spent most of the interaction time coordinating their movement (Technical Coordination) and conversing with each other (Dialogue Management).



(a)

(b)

Figure 5-7: (a) Occurrence-based and (b) time-based analysis of dimensions of collaboration pattern in ASD-NT group and NT-NT group.

5.5 Discussion

We designed and completed a feasibility study of ViRCAS, a novel collaborative activities simulator within CVE. The objectives of the study were to 1) assess the usability of ViRCAS among autistic individuals, 2) assess the ability of the collaborative tasks to support teamwork, 3) observe dimensions of collaboration in the tasks, and 4) examine collaboration patterns in autistic individuals and neurotypical individuals. Our findings offer preliminary support that ViRCAS can assist individuals with and without ASD in learning work-relevant teamwork skills, and capture multimodal data across tasks of varying difficulty levels using different input devices.

The use of multimodal data made it possible to provide quantitative measures of the different dimensions of collaboration. For example, we were able to use input device data to assess technical coordination. The same information may not be easily available from the transcribed speech or other data. Also, multimodal data provided us with quantitative measures of collaboration that human observers might not capture from observing the session or watching a video recording of the interaction such as gazed objects and manipulated objects, which can be difficult to capture by human observers but contain important information to represent collaborative actions. In general, multimodal data analysis can provide higher accuracy compared to unimodal analysis [61]. Currently, the simulator is comprised of three collaborative tasks across varying vocational domains, but it is not limited to these tasks alone. Future work can evaluate additional task types to determine the relevance and performance of the system across different job-relevant teamwork scenarios.

Effective teamwork requires collaborative effort by individuals to work together to achieve a common goal [27]. We embedded 9 dimensions of collaboration that can represent teamwork in our collaborative tasks. Participants in both the ASD- NT group and NT-NT group showed similar patterns of collaboration, which could indicate that the tasks met the universal design principles where participants exhibit similar responses even though they have different abilities. When observing occurrence-based and time-based analysis of the collaboration dimensions, we found that the occurrence frequency of a dimension does not correlate with the duration of the same dimension. For example, for “technical coordination” and “consensus”, the percentage of occurrence for both were about the same in both groups, but when we looked at the duration of these dimensions, participants showed “technical coordination” five times longer than “consensus”. This was because it took more time for the participants to coordinate their movements and only a small fraction of time to show consensus, even though it happened just as frequent. Occurrence-based analysis provides an overall view of dimensions that contributes to the success of collaboration and teamwork, while time-based analysis provides a detailed qualitative evaluation of the dimensions. This comparison is in line with the rating scheme discussed by Meier et al. in [39]. On another note, task division and time management dimensions were less than 1% in both groups for occurrence and time analysis. It is possible that the tasks were designed in a structured manner that offered fewer opportunities for the participants to divide them, and that participants were afforded ample time to not seek time management strategies. In the future, we plan to modify the tasks to provide more opportunities for task division and time management.

The increased in the difficulty levels increased participants' collaboration, as can be seen by the higher number of utterances and back- and forth conversations in the Hard level as compared to the Easy level. We did not compare the utterances in the Tutorials as they focused on task familiarization. We increased the ambiguity and task interdependence as we increased the difficulty level, which motivated participants to ask more questions or describe the task more to each other to proceed with the task. The utterance analysis showed increased collaborative effort for all participants working together. This observation is consistent with other studies that suggest more word usage can influence collaborative learning and learning gains [59]. Paired with dialogue acts analysis, we found that the increase in the number of utterances was related to the tasks. We found statistically significant increases for utterances labeled as Acknowledgement ("Acks"), Describe ("Desc"), Negative sentiment ("Neg"), Questions ("Ques"), "Inform", and "Read", which are all task-oriented conversations. For the Furniture Assembly task, we observed that utterances labeled as "Read" were fewer in the more difficult level because the written instruction manual was removed in the more difficult level, while utterances labeled as Describe ("Desc") increased significantly for all participant because they needed to describe what they see without the instruction. We also found that ASD-NT group used more description ("Desc") utterances compared to the NT-NT group, which could indicate that pairs in the ASD-NT group needed additional explanation and description when performing the task together. NT-NT pairs spoke more to each other than ASD-NT pairs, but that ASD-NT verbal communication increased with task difficulty. In future work, we will analyze in more detail whether those utterances are correlated to improved teamwork and collaboration.

In the gaze analysis, we found that both autistic and NT individuals have similar gaze duration in the PC Assembly and Furniture Assembly tasks. However, for Fulfillment Center, NT participants spent 3 times longer looking at the crates compared to participants with ASD. Because the Fulfillment Center task involves driving a forklift, it is possible that aspects of driving (e.g., familiarity and comfort with driving, ability to focus on driving-relevant stimuli) artificially impacted performance of autistic participants. Studies related to driving in autistic young adults have reported reduced gaze awareness on targeted areas [62], and altered gaze patterns compared to control groups [63], which is consistent with our findings for the Fulfillment Center task. Therefore, future work may consider assessing driving familiarity and comfort when utilizing a task with a driving component. Also, we should consider the calibration error that might influence the accuracy of the gaze results. However, since we are also recording data from the input devices, the data from the input device can improve the accuracy of the eye gaze data.

Our work emphasized the input of stakeholders in preliminary task design and in offering feedback on the developed system. We conducted *t*-tests to examine differences in mean group responses to questions about aspects of the system. We found significant differences for two questions. The first question was related to the comfort level when the task was challenging. Both autistic and NT participants felt comfortable and confident when performing collaborative tasks. However, as the tasks became more challenging, NT participants felt less comfortable than the autistic participants. It is unclear if this was related to the task themselves or to the complexities of social interaction with autistic partners. The second question was related to whether participants would use the system to practice teamwork, if available. We found that autistic participants rated themselves as more likely to use the system than NT participants.

NT participants would not find the system as practical or helpful because they probably do not need the same level of training to practice teamwork skills and daily interactions with others would be enough. A survey reported that ASD individuals had less opportunity to participate in social events/situations as compared to NT individuals [64], thus they see this system as one way to help them practice such skills.

As for the input device preference, the haptic device was the least favorite device compared to the other devices, which indicated that ease of use was more important to the participants than immersive interaction since participants were least familiar with the use of the haptic device. However, existing studies that explored the use of haptic devices in VR-based interactions have shown that with practice, haptic devices can be well accepted by participants [23], [28]. Therefore, for future studies involving teamwork skills training, we will use either the gamepad or keyboard and mouse, while haptic device could be used in studies that examine fine motor skills training.

5.6 Conclusion

Teamwork skills are one of the core skills sought by employers as they can contribute to improved productivity and workplace performance [12]. However, differences in communication and social interaction skills in autistic adults relative to their colleagues can lead to poor teamwork performance, thus limiting employment opportunities for autistic individuals where a high level of teamwork is required [11]. Motivated by this, we designed a novel collaborative activities simulator within CVE, ViRCAS, to support teamwork skills practice for both autistic and neurotypical adults. Results from the feasibility study with 12 participant pairs indicated three main achievements: i) preliminary acceptance of ViRCAS, ii) collaborative tasks that allowed both autistic and neurotypical individuals to communicate and collaborate with each other, and iii) promising potential to quantitatively assess collaboration through multimodal data analysis.

Although the results are promising, it is important to highlight the limitations of the feasibility study and important areas of improvement for future research. First, we had a single-visit study with a relatively small sample size. A longitudinal study with a larger sample size would allow us to examine the impact of training teamwork skills with ViRCAS and enable more complex analyses of the multimodal data. Nonetheless, we believe that these initial results provide justification for an extensive longitudinal study in the future. Second, we did not measure the progress of task performance itself, which could have given us a better understanding of how collaboration affects task performance. To address this, we plan to add a game scoring scheme that can be used to measure task performance for our future study. Third, we did not perform any validation on the eye tracker accuracy and calibration performance that could have impact detection of gaze on smaller virtual objects. Further validation of eye tracking accuracy will be important in future work.

Despite these limitations, results from the feasibility study showed the potential that ViRCAS offers in supporting and nurturing teamwork skills between autistic and neurotypical participants. To our knowledge, this is the first such system and study that investigate the feasibility of a virtual simulator that can support the development and training of teamwork skills for both autistic and neurotypical individuals.

References

[1] [1] M. Guha, “Diagnostic and statistical manual of mental disorders: DSM-5,” Ref. Rev., 2014.

[2] [2] P. M. Dietz, C. E. Rose, D. McArthur, and M. Maenner, “National and state estimates of adults with autism spectrum disorder,” J. Autism Dev. Disord., vol. 50, no. 12, pp. 4258–4266, 2020.

[3] [3] “Growing numbers of young adults on the autism spectrum,” Drexel University Life Course Outcomes. 2019. [Online]. Available: https://drexel.edu/autismoutcomes/blog/overview/2019/June/Growing-numbers-of-young-adults-on-the-autism-spectrum/

[4] [4] X. Wei et al., “Job searching, job duration, and job loss among young adults with autism spectrum disorder,” J. Vocat. Rehabil., vol. 48, no. 1, pp. 1–10, 2018.

[5] [5] D. Hendricks, “Employment and adults with autism spectrum disorders: Challenges and strategies for success,” J. Vocat. Rehabil., vol. 32, no. 2, pp. 125–134, 2010.

[6] [6] A. Karpur, V. Vasudevan, A. Lello, T. W. Frazier, and A. Shih, “Food insecurity in the households of children with autism spectrum disorders and intellectual disabilities in the United States: Analysis of the National Survey of Children’s Health Data 2016–2018,” Autism, vol. 25, no. 8, pp. 2400–2411, 2021.

[7] [7] R. E. Cimera and R. J. Cowan, “The costs of services and employment outcomes achieved by adults with autism in the US,” Autism, vol. 13, no. 3, pp. 285–302, 2009.

[8] [8] C. T. Keating et al., “Autism-related language preferences across the globe: A mixed methods investigation,” PsyArXiv, preprint, May 2022. doi: 10.31234/osf.io/859x3.

[9] [9] S. Baron-Cohen, E. Ashwin, C. Ashwin, T. Tavassoli, and B. Chakrabarti, “Talent in autism: hyper-systemizing, hyper-attention to detail and sensory hypersensitivity,” Philos. Trans. R. Soc. B Biol. Sci., vol. 364, no. 1522, pp. 1377–1383, 2009.

[10] [10] J. C. Kirchner and I. Dziobek, “Toward the successful employment of adults with autism: a first analysis of special interests and factors deemed important for vocational performance,” Scand. J. Child Adolesc. Psychiatry Psychol., vol. 2, no. 2, pp. 77–85, 2013.

[11] [11] M. Waisman-Nitzan, N. Schreuer, and E. Gal, “Person, environment, and occupation characteristics: What predicts work performance of employees with autism?,” Res. Autism Spectr. Disord., vol. 78, p. 101643, 2020.

[12] [12] J. B. Schmutz, L. L. Meier, and T. Manser, “How effective is teamwork really? The relationship between teamwork and performance in healthcare teams: a systematic review and meta-analysis,” BMJ Open, vol. 9, no. 9, p. e028280, 2019.

[13] [13] E. Walsh, J. Holloway, and H. Lydon, “An evaluation of a social skills intervention for adults with autism spectrum disorder and intellectual disabilities preparing for employment in Ireland: A pilot study,” J. Autism Dev. Disord., vol. 48, no. 5, pp. 1727–1741, 2018.

[14] [14] “Neurodiversity hiring: Global diversity and inclusion at Microsoft,” Neurodiversity Hiring | Global Diversity and Inclusion at Microsoft. [Online]. Available: https://www.microsoft.com/en-us/diversity/inside-microsoft/cross-disability/neurodiversityhiring

[15] [15] Specialisterne, “Specialsterne: Assessment.” [Online]. Available: https://www.specialisterneni.com/about-us/assessment/

[16] [16] W. Chen, “Multitouch tabletop technology for people with autism spectrum disorder: A review of the literature,” Procedia Comput. Sci., vol. 14, pp. 198–207, 2012.

[17] [17] V. Bernard-Opitz, N. Sriram, and S. Nakhoda-Sapuan, “Enhancing social problem solving in children with autism and normal children through computer-assisted instruction,” J. Autism Dev. Disord., vol. 31, no. 4, pp. 377–384, 2001.

[18] [18] C. Sung, A. Connor, J. Chen, C.-C. Lin, H.-J. Kuo, and J. Chun, “Development, feasibility, and preliminary efficacy of an employment-related social skills intervention for young adults with high-functioning autism,” Autism, vol. 23, no. 6, pp. 1542–1553, 2019.

[19] [19] E. Salas, N. J. Cooke, and M. A. Rosen, “On teams, teamwork, and team performance: Discoveries and developments,” Hum. Factors, vol. 50, no. 3, pp. 540–547, 2008.

[20] [20] F. Lateef, “Simulation-based learning: Just like the real thing,” J. Emerg. Trauma Shock, vol. 3, no. 4, p. 348, 2010.

[21] [21] S. Parsons and P. Mitchell, “The potential of virtual reality in social skills training for people with autistic spectrum disorders,” J. Intellect. Disabil. Res., vol. 46, no. 5, pp. 430–443, 2002.

[22] [22] L. Bozgeyikli, E. Bozgeyikli, A. Raij, R. Alqasemi, S. Katkoori, and R. Dubey, “Vocational rehabilitation of individuals with autism spectrum disorder with virtual reality,” ACM Trans. Access. Comput. TACCESS, vol. 10, no. 2, pp. 1–25, 2017.

[23] [23] E. Almaguer and S. Yasmin, “A Haptic Virtual Kitchen for the Cognitive Empowerment of Children with Autism Spectrum Disorder,” in International Conference on Human-Computer Interaction, 2019, pp. 137–142.

[24] [24] 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 Trans. Vis. Comput. Graph., vol. 19, no. 4, pp. 711–720, 2013.

[25] [25] Z. Zheng et al., “Impact of robot-mediated interaction system on joint attention skills for children with autism,” in 2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR), 2013, pp. 1–8.

[26] [26] K. Lee, “The Future of Learning and Training in Augmented Reality.,” InSight J. Sch. Teach., vol. 7, pp. 31–42, 2012.

[27] [27] J. E. Driskell, E. Salas, and T. Driskell, “Foundations of teamwork and collaboration.,” Am. Psychol., vol. 73, no. 4, p. 334, 2018.

[28] [28] H. Zhao et al., “INC-Hg: An Intelligent Collaborative Haptic-Gripper Virtual Reality System,” ACM Trans. Access. Comput. TACCESS, vol. 15, no. 1, pp. 1–23, 2022.

[29] [29] 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,” J. Autism Dev. Disord., vol. 43, no. 1, pp. 34–44, 2013.

[30] [30] J. P. Stichter, J. Laffey, K. Galyen, and M. Herzog, “iSocial: Delivering the social competence intervention for adolescents (SCI-A) in a 3D virtual learning environment for youth with high functioning autism,” J. Autism Dev. Disord., vol. 44, no. 2, pp. 417–430, 2014.

[31] [31] F. Vona, S. Silleresi, E. Beccaluva, and F. Garzotto, “Social matchup: Collaborative games in wearable virtual reality for persons with neurodevelopmental disorders,” in Joint International Conference on Serious Games, 2020, pp. 49–65.

[32] [32] S. Sharma et al., “Promoting Joint Attention with Computer Supported Collaboration in Children with Autism,” in Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, San Francisco California USA, Feb. 2016, pp. 1560–1571. doi: 10.1145/2818048.2819930.

[33] [33] A. M. B. Stoit et al., “Internal model deficits impair joint action in children and adolescents with autism spectrum disorders,” Res. Autism Spectr. Disord., vol. 5, no. 4, pp. 1526–1537, Oct. 2011, doi: 10.1016/j.rasd.2011.02.016.

[34] [34] M. Á. Mairena et al., “A full-body interactive videogame used as a tool to foster social initiation conducts in children with Autism Spectrum Disorders,” Res. Autism Spectr. Disord., vol. 67, p. 101438, Nov. 2019, doi: 10.1016/j.rasd.2019.101438.

[35] [35] E. Salas, K. A. Wilson, C. E. Murphy, H. King, and M. Salisbury, “Communicating, coordinating, and cooperating when lives depend on it: tips for teamwork,” Jt. Comm. J. Qual. Patient Saf., vol. 34, no. 6, pp. 333–341, 2008.

[36] [36] S. Wallace, S. Parsons, and A. Bailey, “Self-reported sense of presence and responses to social stimuli by adolescents with autism spectrum disorder in a collaborative virtual reality environment,” J. Intellect. Dev. Disabil., vol. 42, no. 2, pp. 131–141, 2017, doi: 10.3109/13668250.2016.1234032.

[37] [37] S. Parsons, “Learning to work together: Designing a multi-user virtual reality game for social collaboration and perspective-taking for children with autism,” Int. J. Child-Comput. Interact., vol. 6, pp. 28–38, 2015.

[38] [38] Y. Song, L.-P. Morency, and R. Davis, “Multimodal Human Behavior Analysis: Learning Correlation and Interaction across Modalities,” New York, NY, USA, 2012. doi: 10.1145/2388676.2388684.

[39] [39] A. Meier, H. Spada, and N. Rummel, “A rating scheme for assessing the quality of computer-supported collaboration processes,” Int. J. Comput.-Support. Collab. Learn., vol. 2, no. 1, pp. 63–86, 2007.

[40] [40] S. (Suzie) Kardong-Edgren, S. L. Farra, G. Alinier, and H. M. Young, “A Call to Unify Definitions of Virtual Reality,” Clin. Simul. Nurs., vol. 31, pp. 28–34, Jun. 2019, doi: 10.1016/j.ecns.2019.02.006.

[41] [41] S. Martirosov, M. Bureš, and T. Zítka, “Cyber sickness in low-immersive, semi-immersive, and fully immersive virtual reality,” Virtual Real., vol. 26, no. 1, pp. 15–32, Mar. 2022, doi: 10.1007/s10055-021-00507-4.

[42] [42] A. Z. Amat et al., “Collaborative Virtual Environment to Encourage Teamwork in Autistic Adults in Workplace Settings,” in Universal Access in Human-Computer Interaction. Design Methods and User Experience, Cham, 2021, pp. 339–348.

[43] [43] C. Solomon, “Autism and employment: Implications for employers and adults with ASD,” J. Autism Dev. Disord., vol. 50, no. 11, pp. 4209–4217, 2020.

[44] [44] “21st Century Skills and the Workplace.” Gallup. [Online]. Available: https://www.gyli.org/wp-content/uploads/2014/02/21st_century_skills_Gallup.pdf

[45] [45] T. Grandin, “Choosing the Right Job for People with Autism or Asperger’s Syndrome,” INDIANA INSTITUTE ON DISABILITY AND COMMUNITY. 1999. [Online]. Available: https://www.iidc.indiana.edu/irca/articles/choosing-the-right-job-for-people-with-autism-or-aspergers-syndrome.html

[46] [46] M. F. Story, “Maximizing Usability: The Principles of Universal Design,” Assist. Technol., vol. 10, no. 1, pp. 4–12, Jun. 1998, doi: 10.1080/10400435.1998.10131955.

[47] [47] “Logitech Gamepad,” Logitech. [Online]. Available: https://www.logitechg.com/en-us/products/gamepads/f310-gamepad.940-000110.html

[48] [48] “Touch Haptic Device - 3D Systems,” 3d Systems. [Online]. Available: https://www.3dsystems.com/haptics-devices/touch

[49] [49] A. Juliani et al., “Unity: A General Platform for Intelligent Agents.” arXiv, May 06, 2020. Accessed: Dec. 20, 2022. [Online]. Available: http://arxiv.org/abs/1809.02627

[50] [50] “Mirror Networking – Open Source Networking for Unity,” Mirror. [Online]. Available: https://mirror-networking.com/

[51] [51] K. Iiyoshi, M. Tauseef, R. Gebremedhin, V. Gokhale, and M. Eid, “Towards Standardization of Haptic Handshake for Tactile Internet: A WebRTC-Based Implementation,” in

2019 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE), 2019, pp. 1–6. doi: 10.1109/HAVE.2019.8921013.

[52] [52] "Recognize speech from a microphone," Speech to Text – Audio to Text Translation | Microsoft Azure. [Online]. Available: https://docs.microsoft.com/en-us/azure/cognitive-services/speech-service/how-to-recognize-speech?pivots=programming-language-csharp

[53] [53] B. Alibegovic, N. Prljaca, M. Kimmel, and M. Schultalbers, "Speech recognition system for a service robot - a performance evaluation," in 2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV), Shenzhen, China, Dec. 2020, pp. 1171–1176. doi: 10.1109/ICARCV50220.2020.9305342.

[54] [54] A. Jain and P. Bhati, "Comparative Analysis and Development of Voice-based Chatbot System for Differently-abled," J. Phys. Conf. Ser., vol. 2273, no. 1, p. 012003, May 2022, doi: 10.1088/1742-6596/2273/1/012003.

[55] [55] J. Armas, V. Bonifaz Pedreschi, P. Gonzalez, and D. A. Ospina Díaz, "A technological platform using serious game for children with Autism Spectrum Disorder (ASD) in Peru," in Proceedings of the 17th LACCEI International Multi-Conference for Engineering, Education, and Technology: "Industry, Innovation, and Infrastructure for Sustainable Cities and Communities," 2019. doi: 10.18687/LACCEI2019.1.1.278.

[56] [56] "Unity SDK - Tobii Developer Zone," Tobii. [Online]. Available: Unity SDK - Tobii Developer Zone

[57] [57] A. Gibaldi, M. Vanegas, P. J. Bex, and G. Maiello, "Evaluation of the Tobii EyeX Eye tracking controller and Matlab toolkit for research," Behav. Res. Methods, vol. 49, pp. 923–946, 2017.

[58] [58] J. N. Constantino and C. P. Gruber, Social responsiveness scale: SRS-2. Western psychological services Torrance, CA, 2012.

[59] [59] N. Webb, M. Hepple, and Y. Wilks, "Dialogue act classification based on intra-utterance features," Jan. 2005.

[60] [60] J. O'Shea, Z. Bandar, and K. Crockett, "A Multi-classifier Approach to Dialogue Act Classification Using Function Words," in Transactions on Computational Collective Intelligence VII, vol. 7270, N. T. Nguyen, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 119–143. doi: 10.1007/978-3-642-32066-8_6.

[61] [61] A. Mallol-Ragolta, M. Schmitt, A. Baird, N. Cummins, and B. Schuller, "Performance Analysis of Unimodal and Multimodal Models in Valence-Based Empathy Recognition," in 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), 2019, pp. 1–5. doi: 10.1109/FG.2019.8756517.

[62] [62] B. Reimer et al., “Brief report: Examining driving behavior in young adults with high functioning autism spectrum disorders: A pilot study using a driving simulation paradigm,” J. Autism Dev. Disord., vol. 43, no. 9, pp. 2211–2217, 2013.

[63] [63] D. Bian et al., “A Novel Virtual Reality Driving Environment for Autism Intervention,” in Universal Access in Human-Computer Interaction. User and Context Diversity, Berlin, Heidelberg, 2013, pp. 474–483.

[64] [64] C. Lord, J. B. McCauley, L. A. Pepa, M. Huerta, and A. Pickles, “Work, living, and the pursuit of happiness: Vocational and psychosocial outcomes for young adults with autism,” Autism, vol. 24, no. 7, pp. 1691–1703, Oct. 2020, doi: 10.1177/1362361320919246.

CHAPTER 6: MULTIMODAL ANALYTICS OF A USER STUDY WITH THE VIRTUAL TRAINING SIMULATOR FOR IMPROVING TEAMWORK AND EXECUTIVE FUNCTION SKILLS

6.1 Abstract

Teamwork and executive functions (EF) are important skills in the 21st century workforce. Effective teamwork and EF require individuals to collaborate and coordinate with each other which are important learning objectives of teamwork and EF in a training paradigm. However, some individuals with disabilities may experience difficulties working with others due to differences in communication and social interaction skills. We present in this paper the results of a user study with the virtual training simulator with embedded feedback mechanism. This work presents three main contributions: 1) a novel CVE-based virtual teamwork training simulator supporting dyadic interaction with an embedded feedback mechanism; 2) the design of a CVE-based teamwork and EF assessment task using new measures of collaboration adapted from existing literature; and 3) a user study that compares the immediate effect of the training simulator between a training group and a control group. Results of the study showed that the virtual collaborative tasks has the potential to improve teamwork and EF skills of autistic individuals, the feedback mechanism in the training tasks demonstrated positive influence in supporting teamwork interaction between autistic and neurotypical individuals, and the pre-assessment and post-assessment tasks successfully measured quantitative changes in collaborative activities within group and across groups through multimodal data analysis of the dimensions of collaboration.

6.2 Introduction

According to recent studies, teamwork and executive functions (EF) are important skills in the 21st century workforce [1, 2]. Based on a report led by Microsoft Corporation, teamwork skills such as the capability to communicate and collaborate with colleagues and the executive function activities such as time management, planning, and critical thinking (executive functions) are among the core skills sought after in future employees [3]. Studies have shown that teamwork and EF can contribute to improved productivity and workplace performance in a shorter time [4, 5], making the workplace environment to become more collaborative. However, some individuals with disabilities may experience difficulties working with others due to differences in communication and social interaction skills relative to colleagues without disabilities ("neurotypical"). For instance, individuals diagnosed with autism spectrum disorder (ASD) are reported to have reduced communication and social interaction skills needed to work with others [6]. Compared to other individuals with disabilities, adults with ASD have the lowest employment rate, between 20% – 50% [7], and a majority of them with employment are either underemployed or unable to retain their position [8]. Teamwork can help autistic individuals build upon their social communication skills [9], problem-solving skills [10], and self-confidence [11], but they have limited opportunities to practice their communication and collaborative skills with others [12]. These findings highlight the importance of providing autistic adults with opportunities to communicate and work with others to develop their teamwork and EF skills. Although existing training and interventions have shown some improvements in teamwork skills in adolescents with ASD, simulating real-world teamwork scenarios can be tedious, resource-straining, and costly, thus limiting the accessibility and reach of the interventions. Given these circumstances, current intervention and vocational practices to prepare adults with ASD for employment need to be improved.

Human-computer interaction (HCI) technology has shown promising benefits that can potentially complement conventional ASD interventions by providing engaging interactions that can minimize costs with relatively broader access to users [13]. Computer-based commercial games such as Minecraft can be collaborative and have been shown to positively impact changes in teamwork and EF skills [2]. Minecraft is widely used in classroom settings to teach various skills for its engaging appeal, flexibility, and ease of access [2]. However, commercial digital games lack the structure to scaffold skill learning, do not provide real-time feedback or prompts that could facilitate skill learning, and have no objective means of measuring players' skills improvements. A strategically designed collaborative virtual environment (CVE)-based training platform can address the limitations of conventional digital games. First, a CVE-based training platform can be designed with explicit learning objectives that may not be available in digital games. Second, an embedded feedback mechanism in a CVE-based training platform can provide real-time individualized feedback and prompts based on the states of both users and their individual needs [14] to scaffold learning experiences in a collaborative virtual setting [15]. Third, a CVE-based training platform can capture quantitative measures useful for skills assessment. Several studies on CVE-based social communication skills reported improvements in emotion recognition skills and conversational skills in autistic users when they work with others in a series of social scenarios in CVEs [16, 17]. Other CVE-based studies also investigated non-verbal aspects of social communication such as joint attention skills [18], joint action skills [19], and imitation skills [20] where improvements in these skills were observed after completing multiple training sessions in laboratory settings. More recently, researchers found promising results by combining social communication skills and motor skills training for autistic children where they reported increased performance in both domains [21].

Effective teamwork and EF require individuals to work with each other by sharing information, and coordinating their actions, making these as some of the important measures of teamwork and EF in a training paradigm. Recent studies on collaborative learning have identified and defined several dimensions that are relevant for evaluating teamwork and EF [22 - 25]. Meier et al. defined nine dimensions of collaboration in a computer-based collaborative learning environment based on verbal communication that focused on across domain knowledge exchange [25]. Researchers have also defined several non-verbal dimensions involved in collaborative activities for children with ASD [26, 27]. These studies showed that autistic children could provide collaborative responses while playing collaborative computer-based games with another autistic child by observing both their verbal and non-verbal actions. These verbal and non-verbal dimensions of collaboration are important indicators for researchers in designing and measuring teamwork and EF skills.

Feedback mechanism is one of the core learning principles applied in various interventions and training for autistic individuals [28, 29]. In the past, in some instances, feedback was delivered verbally by human observers or therapists during or after the training sessions [30, 31]. In a study by Deitchman et al. [31], the instructor would re-watch the video with the participants and provide feedback on their performance. Manual feedback response can be tedious and as more technology-based interventions are being introduced, researchers have explored the use of automated or embedded feedback mechanisms [32, 33]. White et al. designed a virtual facial emotion expression and recognition training paradigm that incorporated real-time feedback to the participants [32]. Participants reported enjoyment in receiving feedback from the virtual training system. However, autistic individuals have been reported to become dependent on feedback responses during training as discussed by Solomon et al. [34]. To address this issue, researchers have explored the use of a least-to-most feedback mechanism that was based on the participants' dynamic performance [35]. The results indicated that autistic children were able to improve their performance with feedback. However, there has been limited research on feedback mechanisms in a dyadic collaborative interactions based on the performance and behavior of both participants in a dyad.

Previous studies on social communication skills training have discussed the need for an objective, reliable, and cost-effective solution of measuring users' social interactions within the systems [36, 37, 38]. Companies like Microsoft Corporation and Specialisterne, have established neurodiverse hiring programs that include assessment of social and teamwork skills through collaborative tasks, as an alternative to conventional interviews [39, 40]. Such hiring programs will require some standard measures to perceive and evaluate the behavior of the participants. Currently, most of the measurements rely heavily on manual assessments done mainly by human experts and self-reported questionnaires [36, 41-43]. Advancements in artificial intelligence and machine learning technologies over the last two decades have ushered in new methods to quantitatively measure social communication skills and task performance in collaborative virtual interactions using multimodal data [44, 45, 46]. Multimodal analytics is an emerging area of research within HCI that concerns the analysis of integrated data from various sources to identify measurable parameters that can be used to evaluate the skills and provide researchers with an extensive understanding of the skills being learned [47]. Analysis of the multimodal data enabled researchers to observe the dynamic communication patterns and collaboration performance that could complement human assessment and self-reported evaluation [48, 49]. Motivated by the potentials that CVE-based training platform hold in facilitating social skills development and the need to enhance teamwork and EF training experience between autistic and neurotypical individuals, we present in this paper the design and development of an embedded feedback mechanism in a virtual reality-based teamwork training simulator and a collaborative virtual assessment task.

The primary contributions of this work are: 1) a novel CVE-based virtual teamwork training simulator supporting dyadic interaction with an embedded feedback mechanism; 2) the design of a CVE-based teamwork and EF assessment task using new measures of collaboration adapted from existing literature; and 3) a user study that compares the immediate effect of the training simulator between a training group and a control group. The teamwork training simulator consists of collaborative tasks that were carefully designed in our previous work [50]. In this work, we have substantially expanded the teamwork training simulator with a feedback mechanism that provides an individualized response based on the dyadic performance and current state of the dyad. In addition, we have condensed the 9 dimensions of collaboration discussed in [50] into 5 dimensions to better reflect collaborative and EF skills through activity-based collaborative interactions. In order to assess any changes in teamwork and EF skills we have designed, developed, and validated a new assessment task embedded with the dimensions of collaboration and used multimodal data as quantitative measures of teamwork and EF. We conducted a user study to observe any differences in teamwork and EF between a group practicing with the teamwork training simulator (Training group) and a group that did not practice with the training simulator (Control group). The remainder of the paper is organized as follows: Section 2 presents the system design, dimensions of collaboration conceptualization, and the system architecture. Section 3 describes the experimental setup. Section 4 presents the methods taken to analyze the results and the discussion of the results. Finally, Section 5 addresses the potential and limitations of the current study. Relevant literature are cited throughout the paper.

6.3 System Design

This section discusses the design of the collaborative training tasks and the assessment task. First, we present the framework of both training and assessment tasks, followed by the details of the training and assessment task design based on this framework, and finally we described the architecture of the systems.

6.3.1 Design Conceptualization

When designing a computer-based training simulator, the first step is to structure the design to the needs of the target population. Since the focus or learning objective of our simulator is to foster teamwork and EF skills for individuals with deficits in social communication and interaction skills, a dyadic interaction might be more suitable to minimize social anxiety in these population [52]. The other factor we considered in our design was a standardized measure of collaboration through the identification of relevant dimensions of collaboration from previous literature [25]. Nine dimensions of collaboration were selected from previous literature related to dyadic interactions and collaboration [22-25]. These dimensions use both verbal and non-verbal communications that can be quantitatively measured to represent the quality of collaboration between the participants. In a previous feasibility study of the virtual training tasks, we were able to observe these dimensions through multimodal data that we captured and analyzed. During the analysis, we found that these dimensions relied heavily on verbal communication since the study conducted by Meier et al. [25] was observing collaboration in a discussion setting. As such, in our current work, we modified and revised the 9 dimensions of the collaboration into 5 dimensions to be more reflective of activity-based collaboration to match the nature of our collaborative training tasks and collaborative assessment task. Table 6-1 summarizes the modification done to the dimensions of collaboration and the multimodal data used to measure them.

Table 6-1: Dimensions of collaboration and multimodal data mapping

Dimensions	Definition	Multimodal Data	Description
Dialogue Management	• Participants engage in back-and-forth communication and activities.	Utterances - Initiations with responses	Count of initiations with responses
		Utterances - Initiations without responses	Count of initiations without responses
		Utterances – Dialogue acts labeled with 'Acks', 'Pos', or 'Neg'	Utterances categorized as either showing acknowledgement, positive, or negative sentiments
Information Pooling	• Participants share information with each other. • Participant either provide or ask for information • Sustaining mutual understanding is embedded within this	Utterances – Dialogue acts labeled with 'Inform', 'Read', 'Ques'	Utterances that is either asking questions or providing information or reading from instruction

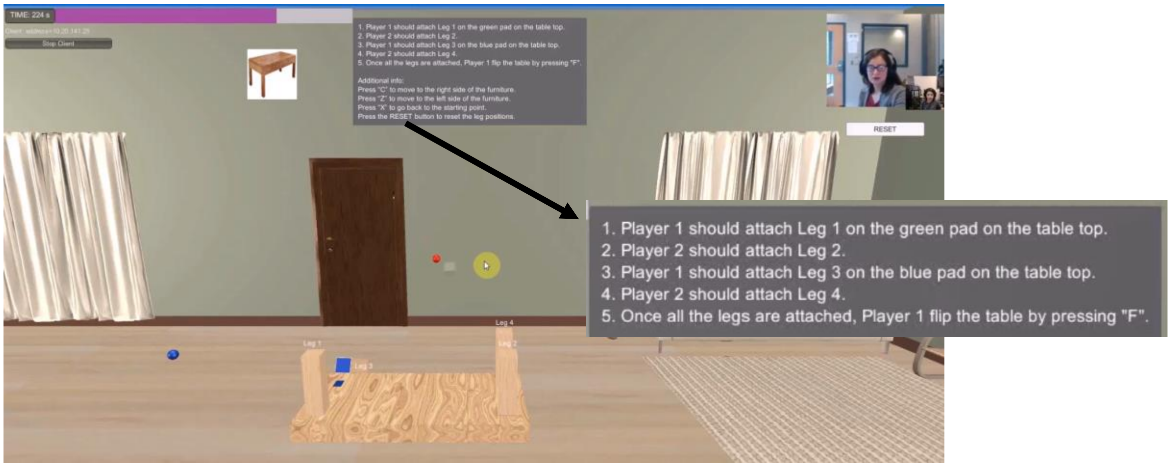
	dimension as showing understanding can happen while exchanging information		
Reciprocal Interaction	• Measure of participants' contribution and how they are distributed. • Individual task orientation is combined with this dimension as contribution to the task is representative of their motivation	Social gaze	Eye gaze detected on video conference window
		Utterances – Dialogue acts labeled with 'Conv'	Utterances that shows positive attitude towards their partner; pleasantries
		Individual scores	Scores received by participant when placing an object at the target
Task division and coordination	• Participants discussed and assigned tasks between them • Participants' ability to maneuver in the task • Reaching consensus and technical coordination are embedded within this dimension as making a decision can happen while dividing the task	Active participation	Instances when participants are actively moving the virtual objects or communicating with their partner
		Task-related gaze	Eye gaze detected on virtual objects
Time management	• Participants display awareness of time constraint and planning	Time-related gaze	Eye gaze detected on the time bar
		Utterances that discuss time constrain in the task	Utterances – 'Time'

6.3.2 Collaborative Training Tasks

We incorporated inputs from stakeholders and end-users in the design process of the collaborative training tasks which included individuals from companies' human resources, certified behavioral interventionist, career counselors and autistic adults. They provided suggestions and feedback throughout the design and development stages of task creation. We also reviewed previous literature related to employment for autistic individuals and in the end, two collaborative tasks were designed as the training tasks, a PC Assembly task [53] and a Furniture Assembly task [54]. These tasks were chosen as they could drive teamwork and EF behaviors between the participants. For each task, we created a tutorial level, an easy level, and a hard level. The five dimensions of collaboration listed in Table 1 were integrated at various instances of the tasks as summarized in Table 6-2. We incorporated three external devices to capture the multimodal data, which were: a) a headset with microphone to capture the speech, b) an eye tracker to capture eye gaze data, and c) a controller to capture button presses and task progression. A prior feasibility study with 6 autistic and neurotypical participant pairs has validated these collaborative tasks abilities to support teamwork skills training for autistic and neurotypical individuals [50].

Table 6-2: Description of collaborative tasks

Collaborative Tasks	Brief Task Description	Enforced dimensions of collaboration
PC Assembly	The objective of the task was for two individuals to work together to attach various virtual computer hardware to build a computer within allocated time. Participants were given installation instructions and access to hardware pieces. They used the keyboard and mouse to select and move the hardware to the correct location.	1. **Dialogue management** – Some installation steps were printed in a different language, enforcing the participants to read out the English instruction to each other as they progress in the task 2. **Task division and coordination** - Participants were given different points of view that limited their view when trying to attach the hardware in place. They had to coordinate their movement to place the hardware correctly. 3. **Information pooling** and **reciprocal interaction**– Each participant was given access to different sets of hardware and installation manuals. They had to exchange installation information and take turns to place the hardware in the correct location. 4. **Time management** – Participants were allocated a limited time to complete the task and thus needed to plan and manage the task within the given time.
Furniture Assembly	Two participants needed to work together to assemble different furniture pieces within the allocated time. They were given written installation instructions and access to the furniture pieces.	1. **Task division and coordination** – In the easy level, the instruction explicitly mentions which participant needs to move which part. For example, “Player 1 attach leg 3 of the table to the blue pad” and “Player 2 attach leg 1 of the table to the white pad on the table”. 2. **Information pooling** – The installation instructions for each participant were different and had missing key information only available to the other participant. 3. **Reciprocal interaction and dialogue management**– Written instruction was not given in the Hard level, only an image of the assembled furniture. Participants had to discuss and agree on a strategy on their own. 4. **Time management** – Participants had limited time to complete the task and needed to plan and divide the task to finish in time



1. Player 1 should attach Leg 1 on the green pad on the table top.
2. Player 2 should attach Leg 2.
3. Player 1 should attach Leg 3 on the blue pad on the table top.
4. Player 2 should attach Leg 4.
5. Once all the legs are attached, Player 1 flip the table by pressing "F".

Additional info:
Press "C" to move to the right side of the furniture.
Press "Z" to move to the left side of the furniture.
Press "X" to go back to the starting point.
Press the RESET button to reset the leg positions.

1. Player 1 should attach Leg 1 on the green pad on the table top.
2. Player 2 should attach Leg 2.
3. Player 1 should attach Leg 3 on the blue pad on the table top.
4. Player 2 should attach Leg 4.
5. Once all the legs are attached, Player 1 flip the table by pressing "F".

Figure 6-1: Example of the installation instruction

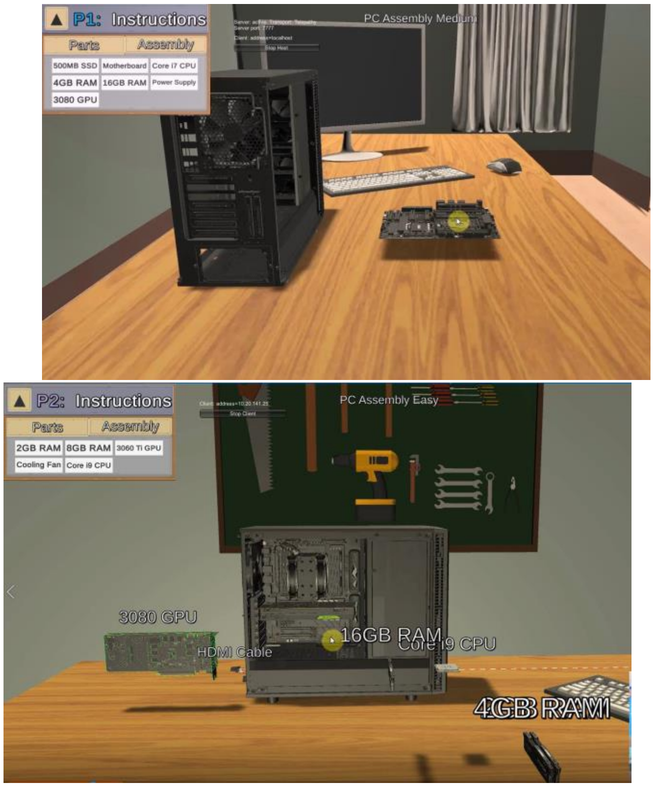


Figure 6-2: The point of view for each participant in the PC Assembly task

6.3.2.1 Feedback Mechanism

The feedback mechanism was designed to provide participants with individualized prompts based on their current performance and behavior in order to support collaborative interaction and foster teamwork and EF skills. We will first briefly explain the participants' behavior labeling before describing the feedback mechanism in more detail.

We developed a rule-based model to label participants' behavior into either *Engaged*, *Struggling*, or *Waiting*, summarized in Table 3. These states were chosen as they were the most common behavior identified in literature related to collaborative interactions [51]. Two annotators trained by a certified behavioral analyst watched video recordings of the participants from the feasibility study [50], and labeled the state of each participant based on the rule-based evaluation as shown in Figure 6-3.

Table 6-3: Collaborative behaviors in dyadic interactions

Behavior	Definition
Engaged	• Participants are interacting with their partner and performing the task. • Measured from transcribed speech and controller input data.
Struggling	• Participants are unable to correctly place an object in the correct location. • Participants are not responding to their partner. • Participants are not focusing their gaze on any object or area of interests. • Measured from transcribed speech, task progression, gaze data, and controller input data.
Waiting	• Participant are not performing any task but still maintain gaze on area of interests. • Measured from task progression, gaze data, and controller input data.

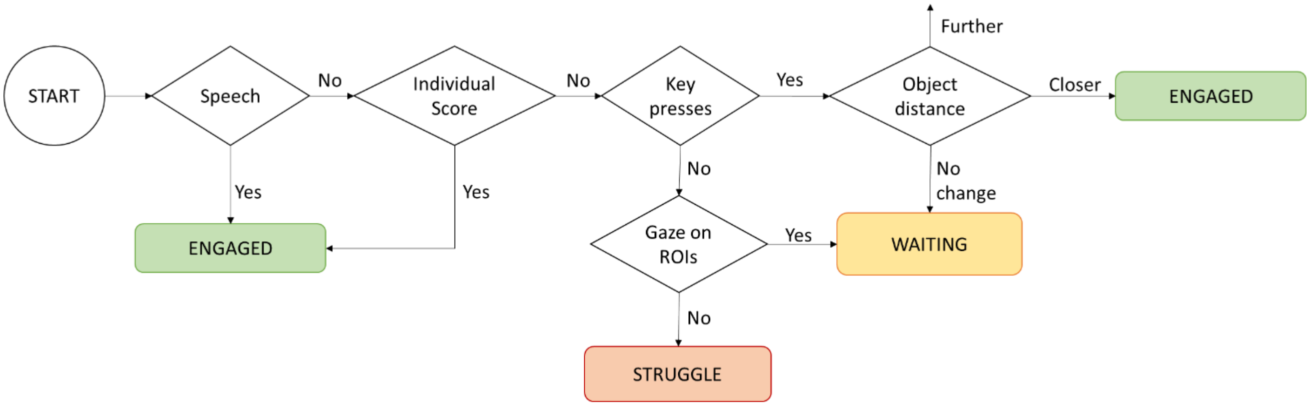


Figure 6-3: Flow-chart for the rule-based evaluation of participants behavior

The feedback mechanism was modeled using a finite state machine (FSM) as illustrated in Figure 6-4. We defined four states in the FSM to represent different aspects of the feedback mechanism. In each state of the FSM, the feedback was designed as a least-to-most prompts, to minimize dependency on the prompts. The initial and default state of the feedback mechanism is *Observe*. In this state, the system actively monitored the participants' performance and behavior. When *Struggling* was detected for more than 30 seconds, the system would transition to a *Help* state. In this state, a prompt would be triggered to both participants. The struggling participant would receive a message to ask them to seek help from their partner, while the other participant would receive a message from the agent to check in with their partner who was struggling. The system would wait in the *Help* state for another 30 seconds, if the participant was showing *Struggling* behavior for another 30 seconds, the system would trigger a similar message to both players more urgently and wait for another 30 seconds. Once the 30 seconds finished, the system would automatically move the piece the participant was struggling with to the correct target location. While in the *Help* state, whenever the participants were no longer *Struggling*, the system would transition back to *Observe* state. A similar logic was implemented for the *Motivate* state. The system would transition into this state when participants were *Waiting* for more than 30 seconds. If participants remained in *Waiting* for more than 90 seconds, the system would prompt them to request help from the researcher. As for the last state, when the system detected a piece was placed correctly, it would transition to the *Positive Feedback* state to provide verbal positive feedback to the participants to continue to motivate them. The system then transitioned back to *Observe* state.

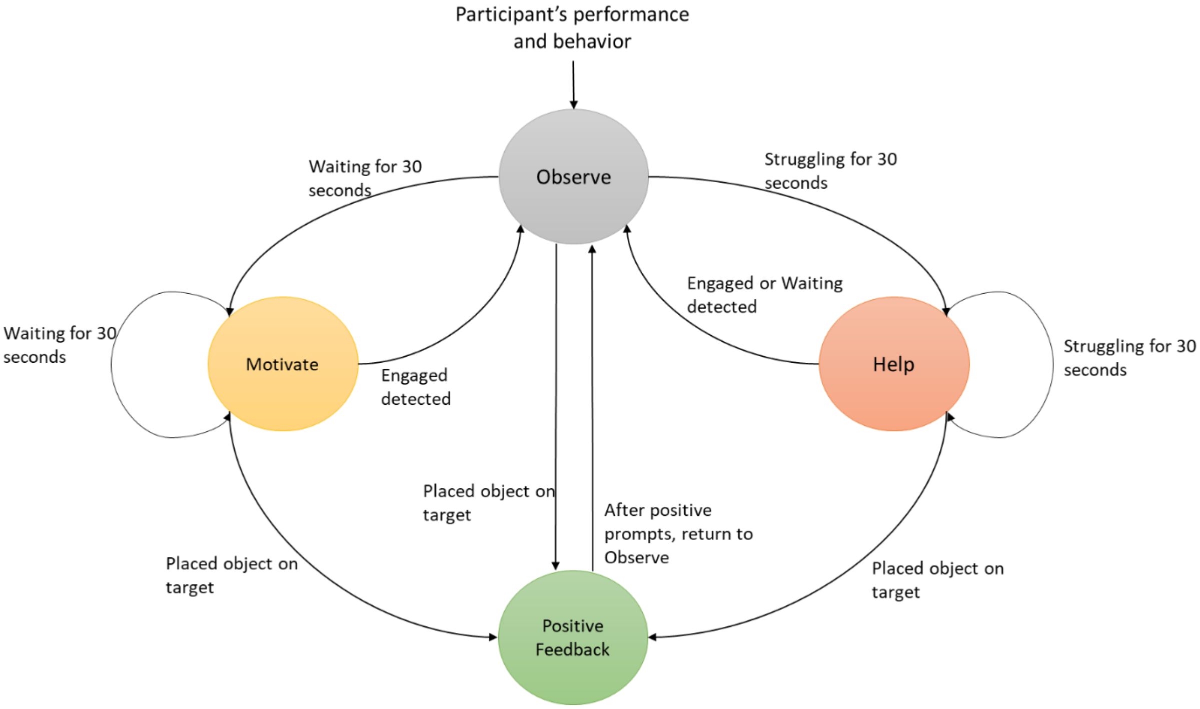


Figure 6-4: Finite state machine of the feedback mechanism

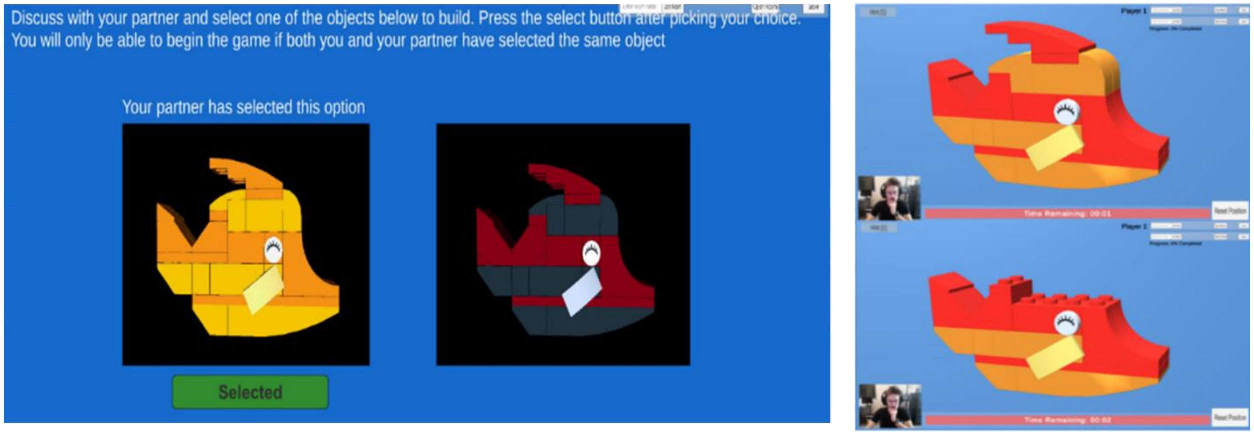
6.3.3 Pre and Post Training Assessment Tasks

Once the training system was in place, there was a need for an assessment module that could be used to test the training system. Thus, we developed pre- and post- training assessment tasks. LEGO based activities were chosen as the assessment tasks because of the potential that LEGO Therapy and its adaptations have shown in collaboration skills training [55-57].

In order to make sure that the design framework remains consistent between the training and assessment tasks, the same five dimensions of collaboration were embedded in the assessment task.

6.3.3.1 Design of assessment tasks

The assessment module included the following levels: i) a Tutorial, ii) Level A, where participants were shown two different variations of an object they need to pick to build as shown in Figure 6-5(a), and iii) Level B where the participants work together to assemble the chosen object. The tutorial was offered to each participant independently to allow them to get acquainted with the controls of the LEGO tasks, while Level A and Level B were collaborative.



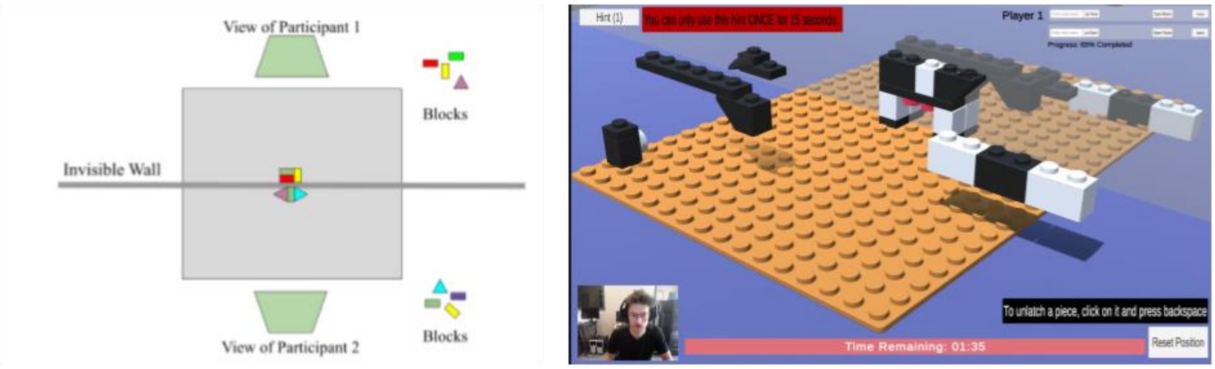
Discuss with your partner and select one of the objects below to build. Press the select button after picking your choice.
You will only be able to begin the game if both you and your partner have selected the same object

Your partner has selected this option

(a) (b)

Figure 6-5: (a) A frame from Level A captured during one of the study sessions;
(b) Map of an object with the ability to go through different layers

At the beginning of Level B, participants were shown the object that they are building for 1 minute. They could go through different layers of the object to look at the structure more closely (see Figure 6-5(b)). In the next step, the participants entered the build mode where they must build the object following what they remember from the map. The participants were across from each other with their virtual environment divided by an invisible wall. The participants were unable to access their partner's side (Figure 6-6). The LEGO pieces were randomly split between the two and each of the participants was required to request access to a specific piece from their partner to build on their side. The participants only gain points if their build looks exactly like the one shown in the map. Both the participants were given one hint each to view the map of the object again for 15 seconds to make room for strategizing together to finish the task in time. Figure 6-7 shows two animals with different configurations that were chosen for pre- and post- tasks to avoid the impact of habituation on the results.



You can only use this hint ONCE for 15 seconds.

Progress: 65% Completed

To unlatch a piece, click on it and press backspace

Time Remaining: 01:35

Figure 6-6: The layout of the build environment (left). A screenshot of the actual build from one of the participant studies (right)

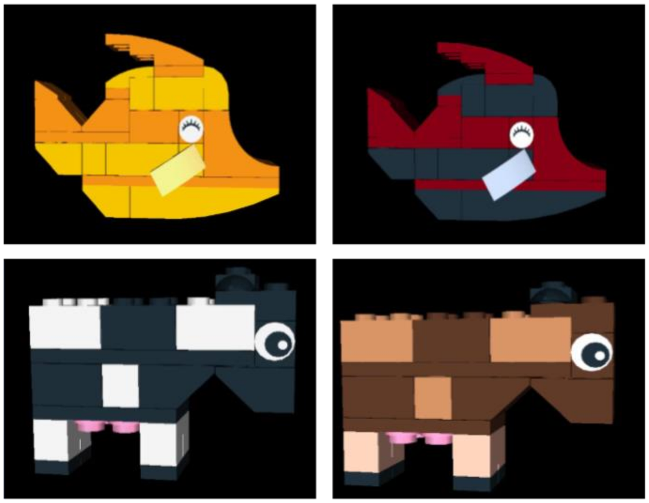


Figure 6-7: Two different variations for the LEGO task for pre-training level (top row) and post-training level (bottom row)

Table 6-4 shows the multimodal data that were collected while the participants were engaged with the task. The difference between data collected during pre- and post- training assessment tasks were used to evaluate the efficacy of the training.

Table 6-4: Data recorded in a log file for both participants using eye gaze, task progression, transcribed speech, and controller input.

MMD	**Description**
Timestamp	Instance of time the data was logged
Label	Player 1 or Player 2
Text	Transcribed Speech
Utterance	Number of words uttered
Duration	Duration of the sentence that was uttered
X/Y_Gaze_Point	X/Y-coordinates of eye gaze on the screen
Focused_Object	Name of the Object the participant was looking at
Total_Score	Percentage of pieces latched by both participants
Individual_Score	Percentage of pieces latched by each participant
Piece_Latched	Name of the piece that was just latched
Piece_Shared	Name of the piece that was just shared
Shared_Count	Number of pieces shared by each player
Brick_Selected	Name of the piece that the user is interacting with
Brick_Select_Duration	Duration of interaction with a selected piece
Active_Effort_Bool	A Boolean indicating active interaction with the system
Map_Interact_Bool	A Boolean indicating interaction with the map
Time_Remaining	Amount of time remaining once the game ends

6.3.4 System Architecture

Unity [58], a commercial game development software, was used to design all the assessment and training tasks in a Collaborative Virtual Environment. For the training system, the setup included a personal computer (PC) with a mouse, a keyboard and a haptic device as controllers. A mouse and a keyboard were used as controllers for the *PC Assembly Task* whereas a haptic device, *Touch* [59], was used as a controller for the *Furniture Assembly Task*. Some key components of digital game design for learning include challenging tasks and immersion [60], that drove the use of different controllers in our training tasks. A webcam and a headset were added to allow for audio-visual (AV) communication. The graphical user interface (GUI) of the game included a window that displayed the video stream of a user’s partner on the screen. This AV interaction allowed to increase the realism of the system and capture important multimodal data (i.e., verbal communication and eye contact). An eye gaze tracker (Tobii EyeX Eye tracker [61]) was used to capture the eye gaze data. Figure xx shows the architecture of the collaborative training and assessment tasks.

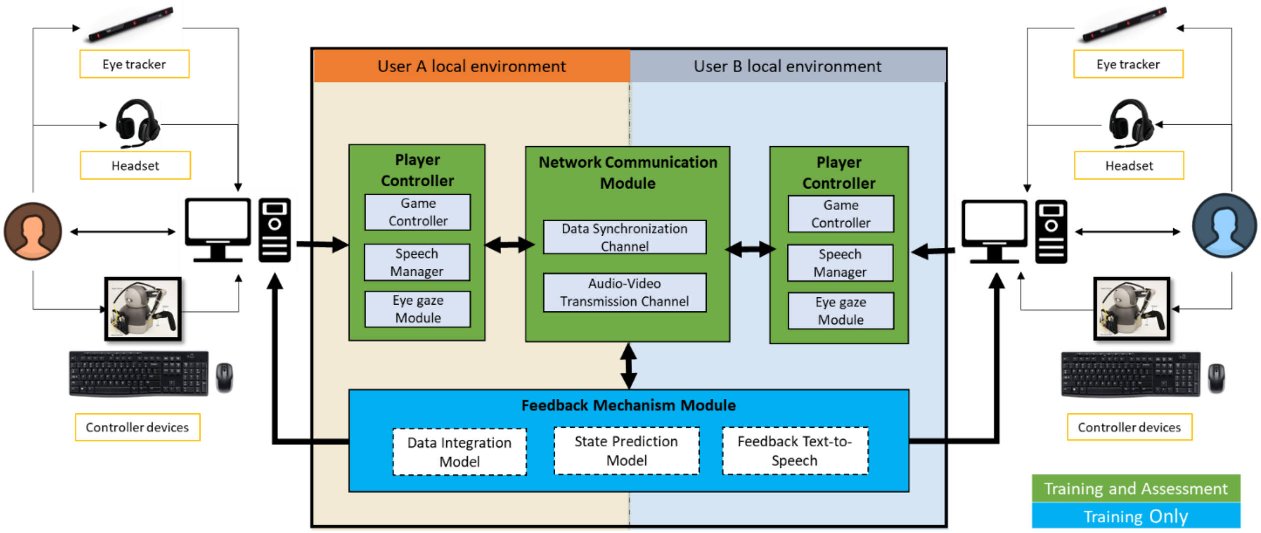


Figure 6-8: System Architecture for the collaborative training and assessment tasks

The network communication module (see Figure 6-8) used *WebRTC API* [62] to allow for AV communication whereas it used *Mirror Networking* [63] for the game environment synchronization between the two users. A feedback mechanism module used the data from game environment as well as transcribed speech that was run through a rule-based state prediction model to detect the state of the user (i.e., W*aiting*, *Struggling*, and *Engaged*). A speech-based feedback was then generated based on the detected state to prompt the users.

As for the assessment tasks, the tasks only used a keyboard and mouse as controllers. In addition, the assessment tasks did not include a feedback mechanism module since the aim of assessment was to allow the users to demonstrate their collaboration skills without any support. The entire network communication layer for assessment task was created using *WEBRTC API* that allowed peer-to-peer AV communication as well as game environment synchronization. An important part of the assessment architecture was the multimodal data logging (shown in Table 4) for post analysis. *WebRTC's DataChannel* was used to transfer the data from Participant 2's PC to Participant 1's PC and store the data of both participants together.

6.4 Experimental Design

A study was conducted to test the efficacy of the proposed training system using the assessment framework. 12 adults with ASD were paired with 12 neurotypical (NT) adults. An ASD-NT pairing was chosen to match real-world circumstances in which individuals with ASD are more likely to end up collaborating with neurotypical individuals in the workplace. Each participant was paired with an individual of the same gender to eliminate the impact of gender biases. Eleven out of 12 pairs identified as males and one pair identified as females within the age range of 17-25 years (mean age= 21.8, SD (Standard Deviation) =2.4). This gender imbalance is consistent with autism gender disparity, i.e., there are significantly more male adults diagnosed with ASD as compared to female adults [64].

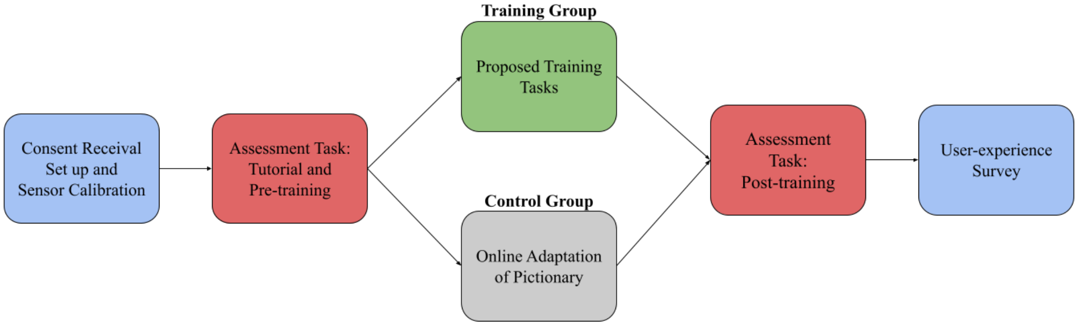

Training Group
Proposed Training Tasks
Consent Receival Set up and Sensor Calibration
Assessment Task: Tutorial and Pre-training
Assessment Task: Post-training
User-experience Survey
Control Group
Online Adaptation of Pictionary

Figure 6-9: Experimental Framework for the study

The 12 pairs are split equally and randomly into two groups: control and training groups. As seen in Figure 6-9, informed written consent is received from each individual (regardless of their group) after which both participants are taken into separate rooms with the PC setup. A headset for audio communication is provided to each participant and the volume is adjusted to the participant's preferred level. The eye tracker was calibrated for each participant after which they go through the tutorial for the assessment (LEGO) task separately. The goal of this tutorial is to get each participant acquainted with the rules, layout, and controls of the assessment tasks. The tutorial is followed by a pre-training assessment task (both Levels A and B) after which the pair either goes through the proposed training or the control activity based on the group they belong to.

Training Group: After the pre-training assessment, the pair goes through three levels -tutorial, easy, and hard - of PC Assembly tasks and three levels of Furniture Assembly tasks. All these levels are equipped with an intelligent agent that uses multimodal data to detect each participant's state (i.e., *Engaged*, *Struggling*, or *Waiting*) in real time and provides the pair with appropriate feedback to nudge them to support each other if a participant is struggling or waiting.

Control Group: After the pre-training assessment, the pair goes through multiple levels of web-based online adaptation of Pictionary, *Drawize* [65], where one participant chose and drew an object while the other partner guessed the drawing and vice versa. This activity was not supported by any audio-verbal communication or multimodal data based real-time feedback that were the key features of the proposed training tasks.

Both the control and training tasks were followed by the post-training assessment tasks (both Level A and B). After each activity, the pre- and post- assessment data (shown in Table 4) were logged for both pairs and saved for post analysis. Table 5 shows the total time allotted for all training, control, and assessment (pre- and post-) tasks. The training activity for the Training group took additional 5 minutes on average for setup and network connection. The post-training assessment task was followed by a survey that was filled in by all the participants to be used for qualitative analysis of user experience. All the experimental protocols were approved by Vanderbilt University's IRB.

Table 6-5: Time Allotted for Training and Assessment Tasks.

Groups	**Pre-training Assessment (minutes:seconds)**		**Training Activity (minutes:seconds)**	**Post-training Assessment (minutes:seconds)**	
	Level A	**Level B**		**Level A**	**Level B**
Control Training	03:00	08:15	15:00	03:00	08:15
Proposed Training			15:00 + 5:00 (Set up time)		

6.5 Results and Discussion

6.5.1 Methods

This subsection describes the steps that were taken to process the multimodal data.

6.5.1.1 Speech Data Analysis

Although the speech data comprises only 10% to 15% of the entire interaction, the information that we could extract from the utterances is useful in various ways to measure the communication aspect of teamwork and EF.

Number of initiations: Two researchers involved in the study established a coding scheme to label the transcribed speech with initiation and responses. They then labeled the log file individually and consolidated any differences through discussion and revision of the labels.

Dialogue acts classification: We wanted to classify the transcribed speech to better understand the context and pattern of the collaborative conversation. In a previous feasibility study [50], we manually labeled the transcribed speech using an existing verbal behavior coding scheme [66] with nine dialogue acts as listed in Table 6-6. The manually labeled data were used to train a natural language processing model, Bidirectional Encoder Representations from Transformer (BERT) [67], that was applied to the data we have in the current study. The trained model achieved an accuracy of 83.1% and an F-1 score of 84%. The confusion matrix is illustrated in Figure 6-10.

Table 6-6: Dialogue acts classification classes

Label	Definition	Example
Acks	Indicate agreement or acknowledge	‘I know’, ‘you’re right’, ‘okay’, ‘yeah’, ‘yup’, ‘cool’, ‘uh-huh’, etc.
Desc	Describe action or intention, decision making	Personal statements of opinion or non-opinion. ‘I think’, ‘I feel’, ‘I believe’, ‘I mean’, etc.
Neg	Disagree, confused, negative statements	‘No I don’t need this one’, ‘I don’t think this is the right one’, ‘no’, ‘um, I’m not sure’, ‘I don’t think so’, ‘oh no’
Pos	Positive feedback from one participant to another.	‘Well done’, ‘good job’
Ques	Questions	‘What do you see?’, ‘can you try W?’
Read	Any indication that the participant is reading task instructions.	Mine says to select the 8 gigabyte RAM'
Inform	Inform, instruct. Action directive statements or statements of instruction from one participant to another.	‘Try moving it more to the right’, ‘and then backwards’, ‘let’s see’, ‘mine has me moving’, ‘let me try’
Conv	Conventional pleasantries	‘thanks’, ‘thank you’, ‘sorry’, ‘my bad’
Out	Uninterpretable. When utterance is incomplete or does not make sense to the coder	‘the. ’, ‘it said an end then snow ’

		Predicted Label						
		Acks	Conv	Inform	Neg	Pos	Ques	Read
True Label	Acks	234	0	30	5	10	4	0
	Conv	0	17	0	0	0	0	0
	Inform	18	0	344	26	6	13	4
	Neg	8	0	18	113	1	4	1
	Pos	1	0	8	1	37	0	0
	Ques	2	0	3	1	0	96	0
	Read	0	0	15	0	0	0	39

Figure 6-10: Confusion matrix for dialogue acts classification model

6.5.1.2 Gaze Data Analysis

We used a TobiiEyeX eye tracker to capture participants’ gaze on the screen while they were performing the task. Before the beginning of the experimental session, we calibrated the participants’ gaze using the Tobii Eye Tracking for Windows application [68]. The gaze data were analyzed into three categories which were:

a. **Task-related gaze**: ROIs and looking at specific virtual objects

b. **Social gaze**: Looking at the video streaming window embedded within the virtual environment

c. **Time-related gaze**: Looking at the timer bar

6.5.1.3 Input Device and Task Progression Analysis

From the input device, we collected the instances when the button was pressed, and the name of the object being manipulated. We used this information to count the number of active actions taken by the participants. As for the task progression, we used four of the parameters logged in Table 6-4 in our analysis:

a. **Overall Score**: Number of pieces attached at the target location, taken from Total_Score.

b. **Individual Score**: Number of pieces each participant attached, taken from Individual_Score.

c. **Active participation**: The accumulated instances when a participant is actively manipulating a piece or speaking to each other, taken from Active_Effort_Bool

d. **Timestamp**: Used to calculate the duration of the task.

Once all the data were processed and consolidated, we calculated the average and standard deviation of the data. We then performed t-tests (paired t-test to compare changes between pre and post, unpaired t-test to compare differences between ASD and NT, and differences between control group and training group)

6.5.2 Overall Task Performance Results

Overall task performance results compared the task score and duration to complete the task. From Table 6-7, participants in both control and training groups showed significant improvement in the total score from pre-test to post-test (Control p-value = 0.02, Training p-value = 0.006). When we looked at the post-test score in more detail, the average score improvement for participants in the training group was higher (41.67) compared to the control group (30). Given the minimal time spent on the training paradigm, the significant improvement in the task score for the training group represents a positive influence of the training paradigm.

Table 6-7: Task performance results

Label	Control		Training		T-test (p-value)	
	Pre	Post	Pre	Post	Control	Training
Overall Score	52.5 (25.45)	82.5 (29.28)	34.2 (14.6)	75.8 (25.8)	**0.0195**	**0.0056**
Duration (minutes)	7:54	7:27	8:00	7:59	**0.0359**	0.1520
Who Asks	P1	P2	P2	P2	n/a	n/a
Who Decides	same	P1	P1	P1	n/a	n/a

We then compared the changes in the individual score between pre-test and post-test tasks for participants in each group as shown in Table 6-8. First, the score increased significantly in the post-test for autistic participants in both the control and training groups, showing that collaborative interaction in a virtual environment can significantly improve task performance for autistic participants. Second, an unpaired t-test showed a statistically significant difference in individual scores between autistic and NT participants in the training group in the pre-test, while in the post-test their scores did not show any significant differences. This observation could indicate that the training paradigms were able to support teamwork practice that allowed ASD and NT pairs to perform at the same level (reciprocity).

Table 6-8: Sub-group performance comparison

Comparison	Group	Individual Score (p-value)
Pre-test to Post-test (paired t-test)	Control Group-ASD	**0.0247**
	Control Group-NT	0.0706
	Training Group-ASD	**0.0069**
	Training Group-NT	**0.0422**
ASD to NT (unpaired t-test)	Pre-test Control Group	0.0856
	Pre-test Training Group	**0.0332**
	Post-test Control Group	0.3466
	Post-test Training Group	0.2127

6.5.3 Utterances Results

Next, we evaluated the back-and-forth interaction between ASD and NT pairs. In terms of dialogue initiations, we found that the number of initiations across all participants was lower in the post-test. This could be a result of habituation where participants became familiar with the task and spent less time in the post-test task. The training paradigm was supported by an embedded feedback mechanism that would ask participants to 'check in' with each other when they were detected as struggling. This has the potential of encouraging and raising awareness in NT participants to continue to engage with their partners as observed in the number of initiations shown by NT participants in the training group. Table 6-9 presents that there was no significant change in the number of initiations by participants in the training group in the post-test, while NT participants in the control group showed a statistically significant decrease in the number of initiations in the post-test. This could mean that NT participants in the control group were less motivated to engage in conversation in the post-test task.

Table 6-9: Number of Initiations

Comparison	**Group**	**T-test**	
		Initiation w/ Resp	**Initiation w/o Resp**
Pre-test to Post-test (paired t-test)	Control Group-ASD	0.2683	0.2762
	Control Group-NT	**0.0204**	**0.0135**
	Training Group-ASD	0.2300	0.0805
	Training Group-NT	0.1123	0.3225

Figure 6-11 illustrates the pattern of dialogue act classification for both the control group and training group in the pre-test and post-test task. In the statistical analysis of the dialogue acts presented in Table 6-10, there were statistically significant increase in the number of utterances providing information and direction ('Inform') and asking questions ('Ques') by the NT participants in the control group. This might indicate that the NT participants in the control group exhibited a leadership role while the autistic participants were the followers.

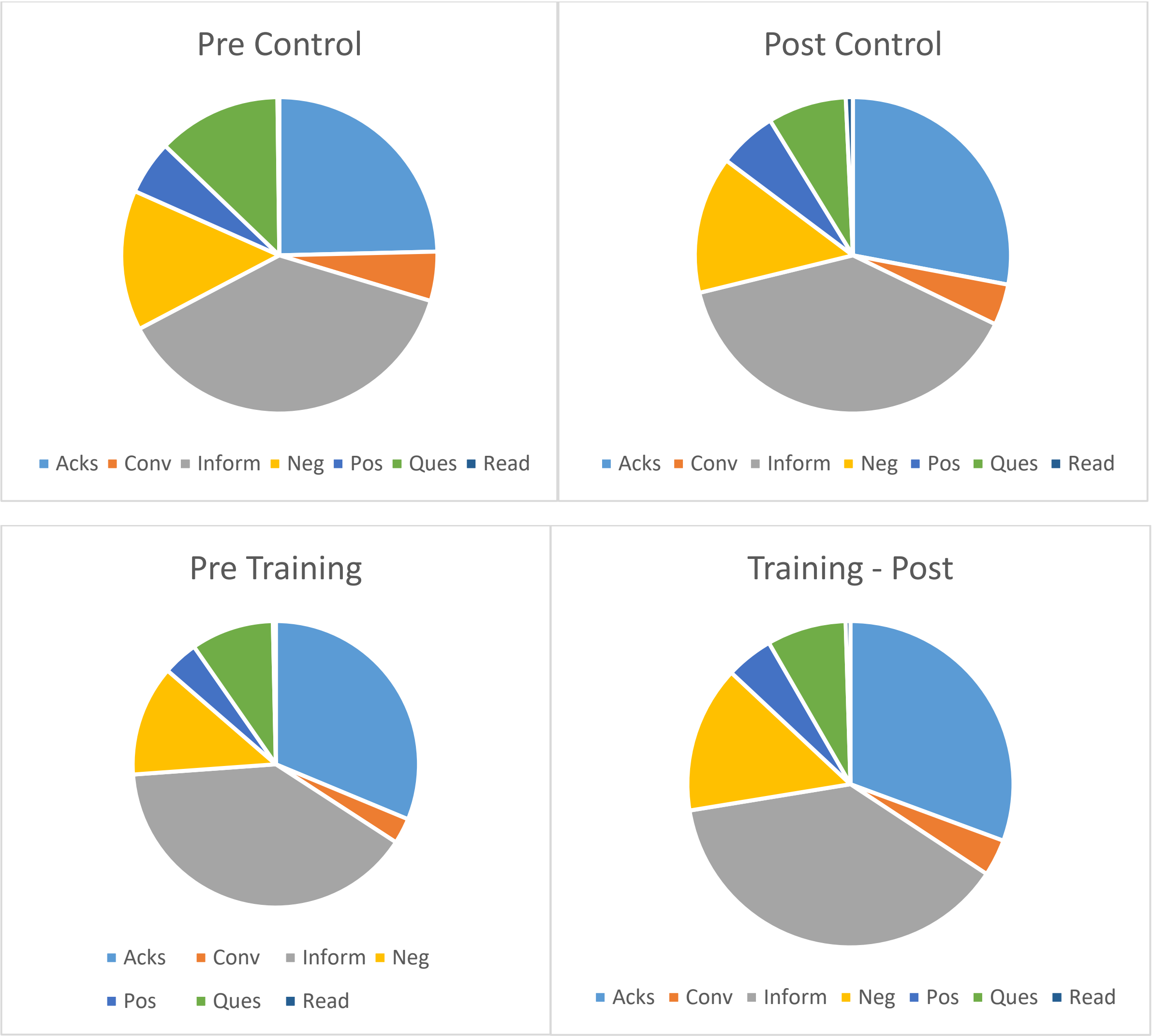


Figure 6-11: Dialogue acts classification for Control group and Training group

Table 6-10: Dialogue acts comparison between pre-test to post-test

Pre-test to Post-test Comparison	Acks	Conv	Inform	Neg	Pos	Ques
Control Group-ASD	0.2816	0.3805	0.0543	0.0997	0.2263	0.1261
Control Group-NT	0.1516	0.0834	**0.0162**	0.1597	0.3032	**0.0038**
Training Group-ASD	0.4221	0.0606	0.1973	0.0812	0.1019	0.1226
Training Group-NT	0.0712	0.1896	0.4984	0.1329	**0.0398**	0.4693

6.5.4 Task Progression Results

The non-verbal data analysis was performed on input controller data, task performance data, and eye gaze data. This analysis provided us with assessment for task coordination and time management skills, which are related to EF skills. First, we looked at the dynamics of the interaction by focusing on the technical effort shown by the participants and how well the pairs distributed the load between them. Active participation was based on the time spent either talking to each other or manipulating the virtual objects such as moving the LEGO pieces to the target location. An unpaired t-test of the pre-test task showed statistically significant difference in active participation between ASD and NT participants in both Control and Training groups, while no significant difference was observed in the post-test task, as seen in Table 6-11. This finding is consistent with the comparison that was made based on the individual score in Table 6-8 that indicated that performing collaborative tasks in a shared virtual environment allowed autistic and neurotypical individuals to contribute equally to the task. Next, we did not find any significant changes to the time management aspect of the interaction. Participants continued to monitor their progress in the post-test task as they did in the pre-test task. Since the post-test task was similar to the pre-test task, participants might not have been too alarmed with the amount of time left, and instead focused their attention on completing the task. This is consistent with the overall results presented previously where participants were able to finish the task in shorter time in the post-test task.

Table 6-11: Active participation comparison

ASD to NT Comparison	Active Effort (p-value)
Pre-test Control Group	**0.0134**
Pre-test Training Group	**0.0361**
Post-test Control Group	0.1497
Post-test Training Group	0.4055

As for the gaze analysis, there were no statistical differences found in the gaze pattern between ASD and NT participants in both groups as shown in Table 6-12. Other studies have claimed that autistic individuals may exhibit atypical gaze patterns when looking at someone. However, in our virtual collaborative environment, we did not observe atypical gaze pattern or avoidance from looking at their partner for autistic participants. It could be that virtual communication was less strenuous and the size of the video conference window was quite small, so it was not overwhelming.

Table 6-12: Gaze fixation on object comparison

ASD to NT Comparison	Task-related Gaze (p-value)	Social Gaze (p-value)	Time-related Gaze (p-value)
Pre-test Control Group	0.1100	0.0603	0.4891
Pre-test Training Group	0.4409	0.1809	0.2431
Post-test Control Group	0.0958	0.0882	0.2447
Post-test Training Group	0.2026	0.1624	0.4607

6.6 Conclusion and Future Work

Teamwork and executive function skills are important skills for employment [2]. Virtual simulation-based training is an effective training method for teamwork and executive function skills that is cost-effective, engaging, and easily scalable for various scenarios [13]. While a feedback mechanism is an important component in a training paradigm to support skills development, existing feedback mechanisms are less effective as they either rely on manual human interventions or performance-based feedback. To address this gap, we designed and developed a teamwork training simulator within a collaborative framework embedded with a feedback mechanism based on both performance and human behavior in dyadic collaborative interaction. We then measured the dimensions of collaboration through a collaborative assessment task. A user study with 12 ASD-NT participant pairs was conducted. Results of the study contributed to these findings: i) the dimensions of collaboration framework in the virtual collaborative tasks showed potential in improving teamwork and EF skills of autistic individuals as can be seen from the improved performance of the autistic participants, ii) the feedback mechanism in the training tasks demonstrated positive influence in supporting teamwork interaction between autistic and neurotypical individuals as seen from the number of back-and-forth communication in the training group, and iii) the pre-assessment and post-assessment tasks successfully measured quantitative changes in collaborative activities within group and across groups through multimodal data analysis of the dimensions of collaboration.

Although encouraging, it is important to mention the limitations of the study and important areas of improvement for future research. First, the short duration of the training time of 20 minutes and a relatively small sample size did not allow us to observe the full extent of the training paradigm. Despite that, we believe that the immediate effect results of the training serves as a motivation for an extensive longitudinal study in the future. A longitudinal study with longer training time and larger sample size would allow us to examine the impact of the training paradigm with feedback mechanism. Second, we did not include explicit instructions on collaborative strategies within the training tasks. Providing participants with clear instructions in specific situations would be beneficial for their learning experience, specifically for autistic participants as it was shown previously that they learn better when they receive direct instructions [69]. For future studies, it would be beneficial to include different collaborative strategies that the participants can incorporate in their training paradigm. Next, it was perceived that the participants were more at ease in the post-assessment task which might be caused by habituation effect. To address this, we would need to revise the post-assessment task to maintain the same framework but with a different activity altogether. Finally, the deterministic design of the rule-based behavior labeling that fed into the feedback mechanism was limited and not adaptable to the dynamic changes of collaborative task. Another future work would involve exploring the use of probabilistic models that are more flexible and improve the robustness of the feedback mechanism.

In spite of these current limitations, we believe that the results presented in this work show the potential of the virtual collaborative framework to contribute to the development and assessment of teamwork and EF skills. To our knowledge, this is the first such study that investigates the training of teamwork and EF skills, and a quantitative method of assessing these skills in dyadic interactions between autistic and neurotypical individuals.

References

[1] Trilling, B., & Fadel, C. (2009). 21st century skills: Learning for life in our times. John Wiley & Sons.

[2] Kulman, R., Slobuski, T., & Seitsinger, R. (2014). Teaching 21st century, executive-functioning, and creativity skills with popular video games and apps. Learning, Education and Games: Volume One: Curricular and Design Considerations, 1, 159.

[3] 21st Century Skills and the Workplace, Microsoft, Pearson Report 2013

[4] Schmutz, J. B., Meier, L. L., & Manser, T. (2019). "How effective is teamwork really? The relationship between teamwork and performance in healthcare teams: a systematic review and meta-analysis." BMJ open, 9(9), e028280. https://doi.org/10.1136/bmjopen-2018-028280

[5] Diamond, A. (2013). Executive functions. Annual review of psychology, 64, 135-168.

[6] Waisman-Nitzan, M., Schreuer, N., & Gal, E. (2020). Person, environment, and occupation characteristics: What predicts work performance of employees with autism?. Research in Autism Spectrum Disorders, 78, 101643.

[7] Hendricks, D. (2010). Employment and adults with autism spectrum disorders: Challenges and strategies for success. Journal of vocational rehabilitation, 32(2), 125-134.

[8] Taylor, J. L., & Seltzer, M. M. (2011). Employment and post-secondary educational activities for young adults with autism spectrum disorders during the transition to adulthood. Journal of autism and developmental disorders, 41(5), 566-574.

[9] Chen, W. (2012). Multitouch tabletop technology for people with autism spectrum disorder: A review of the literature. Procedia Computer Science, 14, 198-207.

[10] Bernard-Opitz V, Sriram N, Nakhoda-Sapuan S. Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. Journal of Autism and Developmental Disorders. 2001;31(4):377-384.

[11] Sung, C., Connor, A., Chen, J., Lin, C. C., Kuo, H. J., & Chun, J. (2019). Development, feasibility, and preliminary efficacy of an employment-related social skills intervention for young adults with high-functioning autism. Autism, 23(6), 1542-1553.

[12] C. Lord, J. B. McCauley, L. A. Pepa, M. Huerta, and A. Pickles, "Work, living, and the pursuit of happiness: Vocational and psychosocial outcomes for young adults with autism," Autism, vol. 24, no. 7, pp. 1691–1703, Oct. 2020, doi: 10.1177/1362361320919246.

[13] Parsons, S., & Mitchell, P. (2002). The potential of virtual reality in social skills training for people with autistic spectrum disorders. Journal of intellectual disability research, 46(5), 430-443.

[14] Kuriakose, Selvia, and Uttama Lahiri. "Design of a physiology-sensitive VR-based social communication platform for children with autism." IEEE Transactions on Neural Systems and Rehabilitation Engineering 25, no. 8 (2016): 1180-1191.

[15] Strickland, Dorothy. "Virtual reality for the treatment of autism." Virtual reality in neuro-psycho-physiology (1997): 81-86.

[16] 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," J. Autism Dev. Disord., vol. 43, no. 1, pp. 34–44, 2013.

[17] J. P. Stichter, J. Laffey, K. Galyen, and M. Herzog, "iSocial: Delivering the social competence intervention for adolescents (SCI-A) in a 3D virtual learning environment for youth with high functioning autism," J. Autism Dev. Disord., vol. 44, no. 2, pp. 417–430, 2014.

[18] S. Sharma et al., "Promoting Joint Attention with Computer Supported Collaboration in Children with Autism," in Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, San Francisco California USA, Feb. 2016, pp. 1560–1571. doi: 10.1145/2818048.2819930.

[19] A. M. B. Stoit et al., "Internal model deficits impair joint action in children and adolescents with autism spectrum disorders," Res. Autism Spectr. Disord., vol. 5, no. 4, pp. 1526–1537, Oct. 2011, doi: 10.1016/j.rasd.2011.02.016.

[20] M. Á. Mairena et al., “A full-body interactive videogame used as a tool to foster social initiation conducts in children with Autism Spectrum Disorders,” Res. Autism Spectr. Disord., vol. 67, p. 101438, Nov. 2019, doi: 10.1016/j.rasd.2019.101438.

[21] H. Zhao et al., “INC-Hg: An Intelligent Collaborative Haptic-Gripper Virtual Reality System,” ACM Trans. Access. Comput. TACCESS, vol. 15, no. 1, pp. 1–23, 2022.

[22] Martinez-Maldonado, R., Gaševic, D., Echeverria, V., Fernandez Nieto, G., Swiecki, Z., & Buckingham Shum, S. (2021). What Do You Mean by Collaboration Analytics? A Conceptual Model. Journal of Learning Analytics, 8(1), 126-153.

[23] Palliya Guruge, C., Oviatt, S., Delir Haghighi, P., & Pritchard, E. (2021, October). Advances in multimodal behavioral analytics for early dementia diagnosis: A review. In Proceedings of the 2021 International Conference on Multimodal Interaction (pp. 328-340).

[24] Crescenzi Lanna, L. (2020). Multimodal Learning Analytics research with young children: A systematic review. British Journal of Educational Technology, 51(5), 1485-1504.

[25] A. Meier, H. Spada, and N. Rummel, "A rating scheme for assessing the quality of computer-supported collaboration processes," International Journal of Computer-Supported Collaborative Learning, vol. 2, no. 1, pp. 63-86, 2007, doi: 10.1007/s11412-006-9005-x

[26] S. Parsons, “Learning to work together: Designing a multi-user virtual reality game for social collaboration and perspective-taking for children with autism,” International Journal of Child-Computer Interaction, vol. 6, pp. 28–38, 2015.

[27] Holt, S., & Yuill, N. (2014). Facilitating other-awareness in low-functioning children with autism and typically-developing preschoolers using dual-control technology. Journal of autism and developmental disorders, 44, 236-248.

[28] Odom, Samuel L., Michelle A. Duda, Suzanne Kucharczyk, Ann W. Cox, and Aaron Stabel. "Applying an implementation science framework for adoption of a comprehensive program for high school students with autism spectrum disorder." Remedial and special education 35, no. 2 (2014): 123-132.

[29] Granpeesheh, Doreen, Jonathan Tarbox, and Dennis R. Dixon. "Applied behavior analytic interventions for children with autism: A description and review of treatment research." Annals of clinical psychiatry 21, no. 3 (2009): 162-173.

[30] Minjarez, Mendy Boettcher, Emma M. Mercier, Sharon E. Williams, and Antonio Y. Hardan. "Impact of pivotal response training group therapy on stress and empowerment in parents of children with autism." Journal of Positive Behavior Interventions 15, no. 2 (2013): 71-78.

[31] Deitchman, C., Reeve, S. A., Reeve, K. F., & Progar, P. R. (2010). Incorporating video feedback into self-management training to promote generalization of social initiations by children with autism. Education and Treatment of Children, 33(3), 475-488.

[32] White, S. W., Abbott, L., Wieckowski, A. T., Capriola-Hall, N. N., Aly, S., & Youssef, A. (2018). Feasibility of automated training for facial emotion expression and recognition in autism. Behavior therapy, 49(6), 881-888.

[33] Bekele, Esubalew, Joshua 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 2016 IEEE virtual reality (VR), pp. 121-130. IEEE, 2016.

[34] Solomon, M., Frank, M. J., Ragland, J. D., Smith, A. C., Niendam, T. A., Lesh, T. A., ... & Carter, C. S. (2015). Feedback-driven trial-by-trial learning in autism spectrum disorders. American Journal of Psychiatry, 172(2), 173-181.

[35] Zheng, Z., Zhao, H., Swanson, A. R., Weitlauf, A. S., Warren, Z. E., & Sarkar, N. (2017). Design, development, and evaluation of a noninvasive autonomous robot-mediated joint attention intervention system for young children with ASD. IEEE transactions on human-machine systems, 48(2), 125-135.

[36] Wang, X., Laffey, J., Xing, • Wanli, Galyen, K., Stichter, J., & Xing, W. (2017). Fostering verbal and non-verbal social interactions in a 3D collaborative virtual learning environment: a case study of youth with Autism Spectrum Disorders learning social competence in iSocial collaborative virtual environment Á Verbal and nonverbal interaction Á Learning environment design. Educational Technology Research and Development, 65, 1015–1039. https://doi.org/10.1007/s11423-017-9512-7

[37] Stichter, J. P., Herzog, M. J., O'Connor, K. V., & Schmidt, C. (2012). A preliminary examination of a general social outcome measure. Assessment for Effective Intervention, 38(1), 40-52.

[38] Hopkins, I. M., Gower, M. W., Perez, T. A., Smith, D. S., Amthor, F. R., Wimsatt, F. C., & Biasini, F. J. (2011). Avatar assistant: improving social skills in students with an ASD through a computer-based intervention. Journal of autism and developmental disorders, 41(11), 1543-1555.

[39] "Specialsterne: Assessment." https://www.specialisterneni.com/about-us/assessment/

[40] "Neurodiversity hiring: Global diversity and inclusion at Microsoft." https://www.microsoft.com/en-us/diversity/inside-microsoft/cross-disability/neurodiversityhiring. (accessed Nov. 02, 2021)

[41] 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, no. 1, pp. 34–44, 2013

[42] Wallace, S., Parsons, S., & Bailey, A. (2017). Self-reported sense of presence and responses to social stimuli by adolescents with ASD in a collaborative virtual reality environment. Journal of Intellectual & Developmental Disability, 42(2), 131-141.

[43] Millen, L., Hawkins, T., Cobb, S., Zancanaro, M., Glover, T., Weiss, P. L., & Gal, E. (2011, June). Collaborative technologies for children with autism. In Proceedings of the 10th international conference on interaction design and children (pp. 246-249).

[44] N. M. Alozie and S. Dhamija, “Automated collaboration assessment using behavioral analytics,” International Society of the Learning Sciences, 2020.

[45] Martinez-Maldonado, R., Gaševic, D., Echeverria, V., Fernandez Nieto, G., Swiecki, Z., & Buckingham Shum, S. (2021). What Do You Mean by Collaboration Analytics? A Conceptual Model. Journal of Learning Analytics, 8(1), 126-153.

[46] Zhang, L., Amat, A. Z., Zhao, H., Swanson, A., Weitlauf, A. S., Warren, Z., & Sarkar, N. (2020). Design of an Intelligent Agent to Measure Collaboration and Verbal-Communication Skills of Children with Autism Spectrum Disorder in Collaborative Puzzle Games. IEEE Transactions on Learning Technologies.

[47] Praharaj, S., Scheffel, M., Drachsler, H., & Specht, M. (2018, September). Multimodal analytics for real-time feedback in co-located collaboration. In European Conference on Technology Enhanced Learning (pp. 187-201). Springer, Cham.

[48] Echeverria, V., Martinez-Maldonado, R., & Buckingham Shum, S. (2019, May). Towards collaboration translucence: Giving meaning to multimodal group data. In Proceedings of the 2019 chi conference on human factors in computing systems (pp. 1-16).

[49] Okada, S., Ohtake, Y., Nakano, Y. I., Hayashi, Y., Huang, H. H., Takase, Y., & Nitta, K. (2016, October). Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets. In Proceedings of the 18th ACM International Conference on Multimodal Interaction (pp. 169-176).

[50] Amat, A. Z., Adiani, D., Tauseef, M., Breen, M., Hunt, S., Swanson, A., Weitlauf, A. S., & Sarkar, N. Design of a Virtual Reality-based Collaborative Activities Simulator (ViRCAS) to Support Teamwork in Workplace Settings for Autistic Adults. IEEE Transactions on Neural Systems and Rehabilitation Engineering [submitted-under revision]

[51] D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. Journal of educational psychology, 105(4), 1082.

[52] D. Spain, J. Sin, K. B. Linder, J. McMahon, and F. Happ´e, “Social anxiety in autism spectrum disorder: A systematic review,” Research in Autism Spectrum Disorders, vol. 52, pp. 51–68, 2018.

[53] C. Solomon, “Autism and employment: Implications for employers and adults with ASD,” J. Autism Dev. Disord., vol. 50, no. 11, pp. 4209–4217, 2020.

[54] T. Grandin, “Choosing the Right Job for People with Autism or Asperger’s Syndrome,” INDIANA INSTITUTE ON DISABILITY AND COMMUNITY. 1999. [Online]. Available: https://www.iidc.indiana.edu/irca/articles/choosing-the-right-job-for-people-with-autism-or-aspergers-syndrome.html

[55] D. B. LeGoff, “Use of lego© as a therapeutic medium for improving social competence,” Journal of autism and developmental disorders, vol. 34, no. 5, pp. 557–571, 2004.

[56] G. Owens, Y. Granader, A. Humphrey, and S. Baron-Cohen, “Lego® therapy and the social use of language programme: An evaluation of two social skills interventions for children with high functioning autism and asperger syndrome,” Journal of autism and developmental disorders, vol. 38, no. 10, pp. 1944–1957, 2008.

[57] S. Didrichsen, “How lego is helping people on the autistic spectrum build a bright future,” Dec 2015. [On-line]. Available: https://www.specialisterneni.com/how-lego-is-helping-people-on-the-autistic-spectrum-build-a-bright-future/

[58] U. Technologies. [Online]. Available: https://unity.com/

[59] https://www.3dsystems.com/haptics-devices/touch

[60] W. L. Bedwell, D. Pavlas, K. Heyne, E. H. Lazzara, and E. Salas, “Toward a taxonomy linking game attributes to learning: An empirical study,” Simulation & Gaming, vol. 43, no. 6, pp. 729–760, 2012.

[61] https://www.tobii.com/products/integration/pc-and-screen-based/tobii-eye-tracker-5l

[62] [Online]. Available: https://webrtc.org/

[63] https://mirror-networking.gitbook.io/docs/

[64] Dietz PM, Rose CE, McArthur D, Maenner M. National and State Estimates of Adults with Autism Spectrum Disorder. Journal of Autism and Developmental Disorders. 2020

[65] https://www.drawize.com/

[66] N. Webb, M. Hepple, and Y. Wilks, “Dialogue act classification based on intra-utterance features,” Jan. 2005.

[67] Alaparthi, S., & Mishra, M. (2020). Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey. arXiv preprint arXiv:2007.01127.

[68] “Unity SDK - Tobii Developer Zone,” Tobii. [Online]. Available: Unity SDK - Tobii Developer Zone

[69] Parsons, S., Leonard, A., & Mitchell, P. (2006). Virtual environments for social skills training: comments from two adolescents with autistic spectrum disorder. Computers & Education, 47(2), 186-206.

CHAPTER 7: A HIDDEN MARKOV MODEL (HMM)-BASED PREDICTION MODEL FOR HUMAN BEHAVIOR RECOGNITION IN DYADIC AUTISTIC-NEUROTYPICAL TEAMWORK TRAINING IN VIRTUAL REALITY

7.1 Abstract

Simulation-based training in virtual environments has been shown to deliver low-cost, engaging, and scalable teamwork training paradigms to intended users, in place of real-world simulation. A real-time feedback mechanism based on human behavior in the training simulator could scaffold learning and promote positive skill growth. Various methods have been explored to recognize or predict human behaviors such as machine learning algorithms, and probabilistic pattern recognition models. However, in a more complex training environment such as for teamwork skills training where dyads must collaborate and interact with each other in the same virtual space, the feedback mechanism would need to look at the changing dynamic of the interactions rather than a static approach. We present the design and evaluation of a rule-based model and a Hidden Markov Model (HMM) to predict three human behavior, which are *Engaged*, *Struggling*, and *Waiting* in collaborative interaction using multimodal data. Validation and performance evaluation showed that HMM achieved higher accuracy across all the selected behaviors and would fit better in a dynamic interaction as it offers more flexibility than a rule-based model.

7.2 Introduction

The ability to work on a team is essential for employment. Teamwork allows for increased productivity in the workplace which in turn maximizes the efficiency of the company. In addition to the benefits teamwork brings to a company, teamwork leads to increased satisfaction in the workplace which can fulfill personal growth. As such, employers seek employees that are effective collaborators. However, individuals with autism spectrum disorder (ASD) may show differences in communication and deficits in social interactions that hinder their ability to find meaningful employment and work well on a team [1]. Of the 5.4 million adults with ASD in the United States, 75% are either unemployed or under-employed relative to their abilities. These perceived deficits in teamwork and collaboration prevent companies from accessing a large talent pool. Therefore, teamwork training is essential for young adults with ASD to prepare them for employment. Nonetheless, it is difficult to create real-life opportunities to practice teamwork-relevant skills. Simulation-based training in virtual environments has been shown to deliver low-cost, engaging, and scalable teamwork training paradigms to intended users, in place of real-world simulation [2]. In our previous work, we developed a series of collaborative tasks using a collaborative virtual environment (CVE) as a teamwork training simulator [3]. A real-time feedback mechanism based on human behavior in the training simulator can scaffold learning and promote positive skill growth. To achieve this, the system would require a reliable way of recognizing human behavior.

Researchers agree that human behavior recognition in feedback mechanisms could lead to a better training experience and improved skill development. For example, studies have reported that using both performance and human behavior in adaptive training paradigms was able to keep participants engaged and improve their training outcomes [4, 5, 6]. However, these studies were limited to single-user interaction. In addition, the current solutions for behavior recognition models rely heavily on human experts to evaluate human behavior. Manual evaluation of human behavior can be resource-straining and prone to bias [7]. Therefore, we propose the development of two mathematical models to automate the recognition of human behavior of two individuals working on virtual team-building tasks.

This paper presents the design and evaluation of a rule-based model and a Hidden Markov Model (HMM) to predict human behavior in collaborative interaction using multimodal data. The primary contributions of this paper include i) a novel dataset that includes multimodal data labeled by expert annotators, ii) the design of a rule-based prediction model for human behavior recognition, iii) the development of a Hidden Markov Model (HMM) prediction model, and iv) the comparison of the performance of the prediction models against hand-labeled data.

7.3 Related Works

The most common and early method of human behavior recognition is manual labeling by human observers. Observers that are involved in manually labeling experimental sessions are usually experts in psychology or human behavioral understanding. Observers either label human behavior in real-time during the experimental session [8] or by watching video recordings of the sessions [9, 10]. More recently, modern video annotation software has been used to complement human observers labeling, where researchers need to provide detailed coding schemes to achieve good coding results [11, 12]. Although manual labeling is proven to be reliable, labels are not available instantaneously, the process can be resource-straining, and the generalizability of the coding scheme is low [13]. To address these limitations, researchers have explored various classification methods to predict human behavior in computer-based interactions [14, 25].

One method includes using a rule-based model where a set of rules were created from existing observations [15, 16]. For example, researchers designed a rule-based fuzzy logic model to predict human emotions using speech [17]. In the study, the rules were constructed based on a collection of utterances that were labeled as ground truth with various speech features and three emotion components. The model achieved an average agreement of 60%-80%, which can be considered a good prediction accuracy given that it was estimating human emotion using only speech information. In another study, researchers designed a more complex rule-based system to evaluate human behavior in an assisted living environment [18]. The model achieved more than 92% accuracy in predicting daily activities such as meal preparation, personal hygiene, and using the toilet. Rule-based models has been shown to have high performance when the rules were clearly distinguished from each other. However, in an application that has a wide range of variations, such as in natural language processing, the rules can become too complex and laborious to create [19]. Additionally, subtle and ambiguous changes in the data may cause the rules to grow and become harder to manage [19].

Other researchers have looked at prediction models that can be more robust and flexible to subtle changes in data, which included statistical machine learning methods [20, 21,22]. One study combined the use of deep learning and rule-based methods to predict engagement and disengagement in a multi-person human-robot interaction paradigm [20]. Deep learning was used to extract engagement-related features from the human participants such as gaze, head pose, and body posture, while rule-based was used in selecting the subject and the engagement decision-making algorithm. The design achieved a 93% F1-score. In another study, Okada et al. used verbal and non-verbal measures to assess the communication skills of individuals in different types of discussion tasks [21]. They captured data from speech and head movement and extracted verbal and non-verbal features from the data. The features were analyzed using eight regression models with different feature combinations and compared the results against human-coded evaluations of communication skills. In a study by Cuayáhuitl et al., they applied a deep reinforcement learning (DRL) method that could play a strategic board game with human users and negotiate with other players [22]. The DRL performed better than other machine learning methods. However traditional machine learning methods are deterministic where the output would be the same for the same input, which may not be representative of certain human behaviors such as affective state.

Alternatively, probabilistic models such as Hidden Markov Model (HMM) have been widely used to model human behaviors [23, 24]. Mihoub et al. presented a probabilistic modeling framework using Incremental Discrete Hidden Markov Model (IDHMM) to recognize and generate multimodal joint actions in a face-to-face interaction [23]. The result indicated that IDHMM produced a higher classification rate than Support Vector Machines (SVM), with a mean cognitive state recognition rate of 92% compared to 81%. Another study implemented double-layer HMM to evaluate individual and group activities in group meetings [24]. The model was tested with 59 public corpora of meeting data and the result showed that the two layers HMM had a 70.3% accuracy, while a single-layer HMM only had 57.5% accuracy. A HMM also offers the added advantage of containing temporal and sequential information in the probabilistic functions, making it a suitable solution as a prediction model of collaborative human behavior in a teamwork training simulator.

7.4 Experimental Design

We conducted a preliminary study to i) gather multimodal data from participants interacting collaboratively with each other, ii) label the data based on defined collaborative behavior, and iii) perform statistical analysis that compares the performance of a rule-based prediction model and a HMM prediction model in recognizing human behavior in dyadic collaborative interactions.

7.4.1 Collaborative Tasks Description

The tasks were designed based on input from stakeholders to encourage teamwork in a workplace environment between an autistic individual and a neurotypical (non-autistic) partner, which was discussed in detail in our previous work [3]. Multiple discussion sessions with the stakeholders were conducted to select tasks that are collaborative and include interactions that are suitable in a workplace environment. The collaborative tasks selection was driven by employment-related studies for autistic individuals. The first task was a *PC Assembly Task* where two participants were located on opposite ends of a table in the virtual environment giving them different views of the workspace. They both were given written instructions and different hardware to collaboratively build a single computer. They would use the keyboard and mouse to move the components into the correct location within a set amount of time. Participants were required to take turns and communicate with each other when assembling the PC. The next task was a *Furniture Assembly Task* where participants were placed in a virtual living room and worked together to assemble various furniture pieces within a set amount of time. They used a haptic device to move the furniture parts to the target area. The final task was a *Fulfillment Center Task* where participants would drive forklifts with varying height capacities to transport crates from a warehouse to a drop-off location. Participants used a gamepad to drive the forklift in this task. Three design strategies were embedded within the tasks to encourage communication and collaboration between the participants: a) incomplete installation instructions were given to each participant to encourage them to exchange information to progress in the task; b) participants were given only an image of an assembled furniture, without written instruction, to encourage them to divide the task and coordinate their actions; and c) components that were only available to one participant but not the other and varying the location of the crate to allow participants to practice turn-taking. These collaborative tasks, as illustrated in Figure 7-1, were designed in Unity, a multi-platform game development software [26].

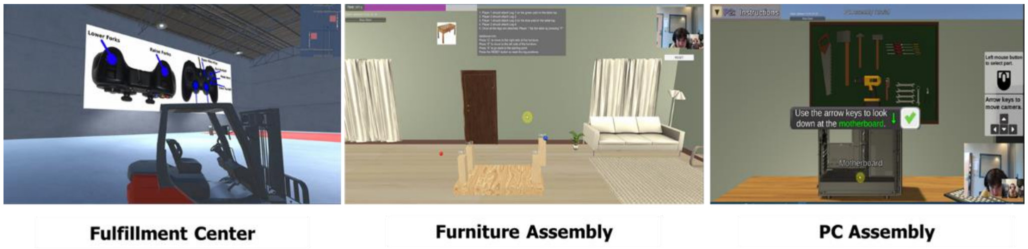

Lower Forks
Raise Forks
Instructions
Left mouse button to select part.
Arrow keys to move camera.
Use the arrow keys to look down at the motherboard.
Motherboard
Fulfillment Center
Furniture Assembly
PC Assembly

Figure 7-1: Collaborative tasks to support collaborative interaction between autistic individuals and neurotypical partners

7.4.2 Participants and Protocol

We recruited 6 autistic participants and 6 neurotypical (NT) participants to form 6 ASD-NT participant pairs. The demographics for the participants are shown in Table 7-1. Participants with ASD were recruited through an existing university-based clinical research registry and the NT participants were recruited from the local community through regional advertisement.

Table 7-1: Participants demographic information

Participants	ASD (N = 6) Mean (SD)	NT (N = 6) Mean (SD)
Age	20.5 (2.8)	22.8 (3.6)
Gender (% male)	50%	50%
Race (% White, % African American)	100%	83%, 0%
Ethnicity (% Hispanic)	0%	17%

Two computers were set up in two separate rooms and used local area network (LAN) to connect them to the same virtual environment. Participants filled out the consent forms before going to the separate rooms. The session lasted about 90 minutes. All study procedures were approved by the Vanderbilt University Institutional Review Board (IRB) with associated procedures for informed assent and consent. Figure 7-2 illustrates the setup of the experiment.

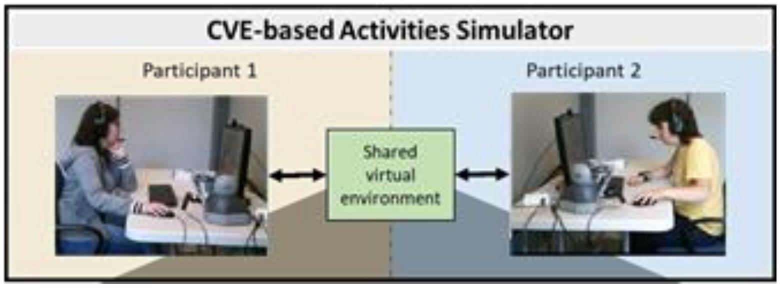


Figure 7-2: System setup where two users in separate rooms perform virtual tasks together

7.5 Methodology

We described four main processes involved in designing, training, and evaluating human states prediction models in collaborative interactions, as seen in Figure 7-3. The multimodal data that were captured from the participants together with video recording were used by annotators to label the participants' states to establish a ground truth. These labeled data were used to design and train a rule-based prediction model and a HMM. We then evaluated both prediction models' performances. The following subsections explain in detail each step of the process.

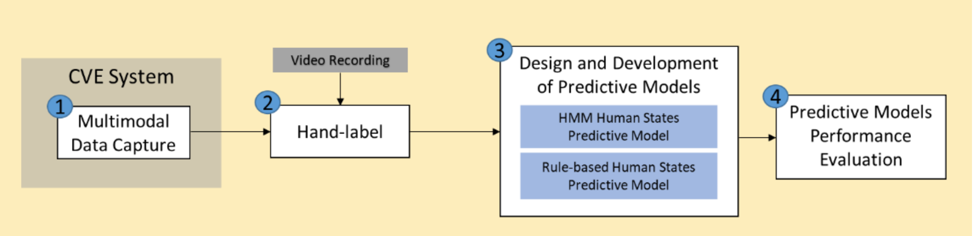

CVE System
1
Multimodal Data Capture
Video Recording
2
Hand-label
3
Design and Development of Predictive Models
HMM Human States Predictive Model
Rule-based Human States Predictive Model
4
Predictive Models Performance Evaluation

Figure 7-3: Workflow for prediction models design and evaluation

7.5.1 Collaborative Behavior Coding Scheme

A study on collaborative learning reported that the most frequent behavior experienced by participants when working collaboratively were engagement, confusion, and boredom [27]. Based on this literature review and discussions with the stakeholders and behavioral analysts, we chose three behaviors that would be the most useful to recognize in our teamwork training simulator; *Engaged*, *Struggling*, and *Waiting*. These three behaviors represent essential stages of teamwork allowing the system to provide informed and meaningful feedback. *Engaged* captures the state when the participant is performing the task and collaborating with their partner [28], allowing the system to provide positive praises such as 'Good job!' or 'Keep up the good work!'. *Struggling* represents the state when the participant were not progressing in the task (e.g., the object was moving away from the target), not communicating for a while with their partners, or was disengaged with the task (e.g., looking outside the focus area for some time) [29]. The system would then use the *Struggling* state as an indicator to prompt the participants to help each other, for example 'Ask your partner to help you with the task' and to the other participant 'Your partner seems to be struggling, offer them help'. Turn-taking is part of teamwork and collaborative interaction, and we are using *Waiting* to capture when the participant is waiting for their partner to perform a task [30] and differentiate it from when a participant is distracted or disinterested (which is categorized under *Struggling*). In *Waiting* state, the system would allocate some time for the participants to wait without prompting the participants. Although there are only three states discussed in this work, more states could be added in the future based on the need and understanding of collaboration and teamwork. A coding scheme was defined in consultation with a certified behavioral analyst, to ensure the consistency of the manual labeling, shown in Table 7-2.

Table 7-2: Definition of Participant States

#	Participant State	Definition
1	Engaged	The user is focused on the task and progressing well without struggling – either talking to their partner or not, looking at specific ROIs or focus area, controller usage detected, object moving closer to the target, successful attempts detected
2	Struggling	The user shows signs of struggling to progress in the task – not talking to their partner, looking at specific ROIs or outside, repetitively using the controller, object not moving or moving away from the target, no successful attempts detected
3	Waiting	The user is waiting for their partner; no speech detected, looking at specific ROIs or focus area, not using any controllers, objects (moved by partner) getting closer to the target, successful attempts (by their partner) detected

7.5.2 Multimodal Data Capture

There were three devices for the multimodal data capture in the collaborative system. We used a game controller, a microphone headset, and an eye tracker for each participant to derive seven binary features that could represent the current state of the participants in collaborative interactions. The diagram in Figure 7-4 shows that we derived four binary features from the game controller, one binary feature from the transcribed speech, and one binary feature from the eye tracker.

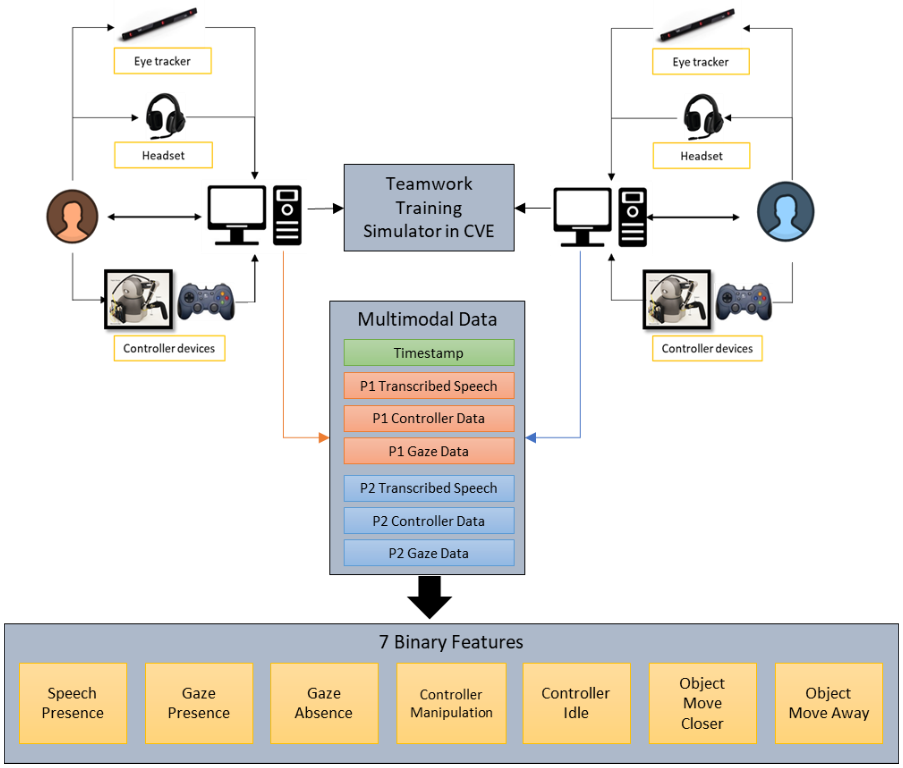


Figure 7-4: Feature extraction from three peripheral input devices

First, we needed verbal data, which we gathered from the transcribed speech. The Speech Presence feature was mapped to '1' when there was the presence of speech and '0' otherwise. Then, we wanted information on task participation and activities within the virtual environment. We use the data from the controller to capture the presence of controller keypress and the distance of the object from the target. The Controller Manipulation feature is recorded as '1' when a keypress was detected, otherwise the feature was set to '0'. The opposite rule was applied to the Controller Idle feature. For the Object Move Closer and Object Move Away features, the binary values were determined by the distance of the object from the target. The Object Move Closer feature was recorded as '1' when the object moved closer to the target and '0' when the object was not, while the Object Move Away feature was recorded as '1' when the object moved further away from the target and '0' when the object was not. When the feature values were both '0', it would mean that the object was not moving.

For non-verbal communication, we are capturing the eye gaze data. We extracted two gaze features based on the region of interest (ROIs) and focused gaze points. The Gaze Presence feature was recorded as '1', and the Gaze Absence feature was recorded as '0' when a ROI was detected or the gaze points were within the defined 'focus area' as depicted in Figure 3. When there was no ROI registered or the gaze points were outside the 'focus area', then the Gaze Absence feature was recorded as '1' and the Gaze Presence feature was recorded as '0'.

These binary features were concatenated to form a feature vector. All of the features were collected continuously with a sampling rate of 1 sample per second. The binary values and feature vectors that were extracted from the Pilot Study were used to design the rule-based model and HMM, respectively.

7.5.3 Hand Labeling to Establish Ground Truth

In the next step, two annotators, trained by a certified behavioral therapist, used the coding scheme described in Table 7-2 to label the participants' states into either '*Engaged'*, *Struggling'*, or '*Waiting'* based on the extracted features and video recording of the sessions. The annotators labeled 10 minutes worth of interactions from each session. Both annotators label all six sessions of data separately using the established scheme and achieved 98% agreement. The 2% disagreement was reconciled when both annotators decided on an agreed label through discussion. Of the four labeled sessions, the class distributions of the three states were as follows: *Engaged* - 19.9%, *Struggling* - 28.0%, and *Waiting* - 52.1%.

7.5.4 Rule-based Prediction Model Design

We designed a rule-based prediction model for our collaborative tasks to provide real-time state prediction for each participant when they are collaborating, replicating the role of human annotators. Based on the coding scheme in Table 7-2, we created step-by-step rules based on the feature values to predict the participants' states. We then consolidated these rules into a flow chart depicted in Figure 5.

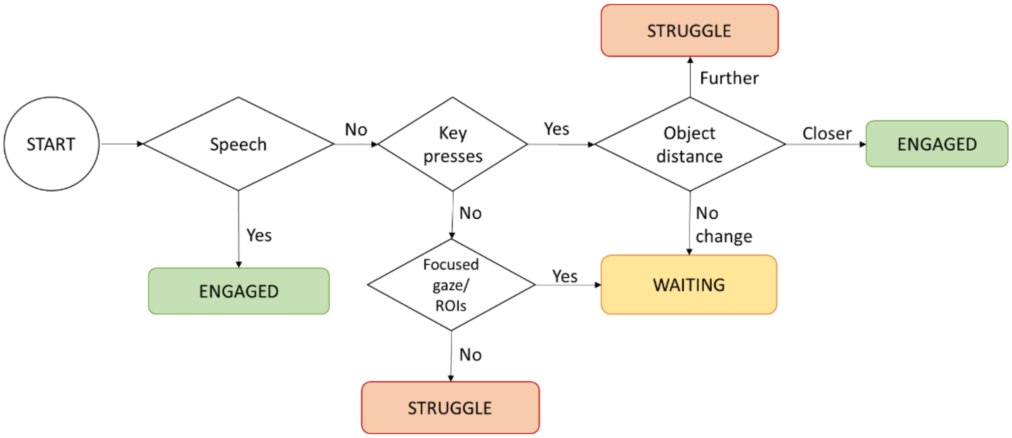


Figure 7-5: Flowchart for the rule-based prediction model

7.5.5 HMM Design and Training

Probabilistic predictive models offer flexibility and scalability compared to deterministic predictive models [31]. We defined the five main elements of a HMM [32] in Table 7-3. An ergodic state transition model was designed for our model as we assumed that the collaboration state can change from one state to any of the other states. Figure 7-6 shows a possible diagram of the HMM.

Table 7-3: Definition of HMM elements

Symbol	Definition	Values
N	Number of hidden states in the model	$\{Engaged, Struggling, Waiting\}$
M	Number of distinct observations	We are using a 7-digit binary vector based on the extracted features from the multimodal data $\{Obs_1, Obs_2, Obs_3, \dots, Obs_M\}$ Example values: 1101010, 0010100
A	State transition probability distribution - Probability matrix of transition from one state to another.	Matrix size is NxN, in our case 3x3. Values of the matrix are generated from training the model $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$
B	Emission probability distribution - Probability matrix of observing a particular observation in the current state	Matrix size is NxM. Values of the matrix are generated from training the model $\begin{bmatrix} b_{11} & \cdots & b_{1M} \\ \vdots & \ddots & \vdots \\ b_{31} & \cdots & b_{3M} \end{bmatrix}$
π	Initial state probability distribution	Initial state probability matrix, usually equally distributed $[0.3 \quad 0.3 \quad 0.3]$

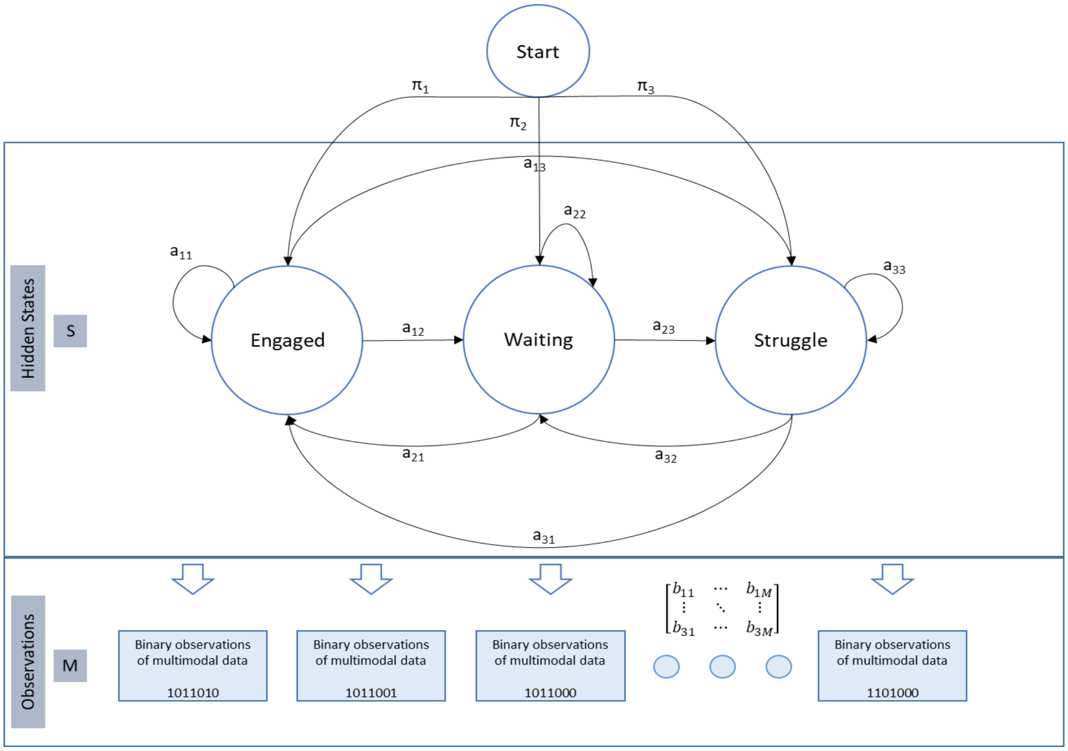

Start
π_1
π_2
π_3
a_{13}
a_{22}
a_{11}
a_{33}
Hidden States
S
Engaged
Waiting
Struggle
a_{12}
a_{23}
a_{21}
a_{32}
a_{31}
Observations
M
Binary observations of multimodal data 1011010
Binary observations of multimodal data 1011001
Binary observations of multimodal data 1011000
$\begin{bmatrix} b_{11} & \cdots & b_{1M} \\ \vdots & \ddots & \vdots \\ b_{31} & \cdots & b_{3M} \end{bmatrix}$
Binary observations of multimodal data 1101000

Figure 7-6: HMM diagram of all elements

We consolidated the hand-labeled behavior and multimodal data to design and train a HMM in Matlab [33]. Using the Statistics and Machine Learning Toolbox [34] available in Matlab, we utilized the functions available to generate an initial HMM, then train the model to achieve optimal performance. First, we entered the sequence of binary vectors as the observation sequence and the hand-labeled states as the hidden states. To optimize the HMM training, we introduced a *k*-fold cross-validation method where we divided the dataset into *k* number of folds. Then we split the data into 70-30 ratio, where 70% of the data were used as training data and 30% of the data were used as testing data. Figure 7-7 represents the *k*-fold cross-validation and training-testing data split. Next, we used the **hmmestimate** function to generate estimated transition and emission metrices for the model by calculating the maximum likelihood using 70% of the observations and hand-labeled states. Once we have the estimated transition and emission metrices, we used the Matlab function **hmmtrain** to predict the participants' states using the 30% observation sequence allocated for testing and compared the results to the hand-labeled states. The entire process of estimating and predicting the HMM were repeated *k* times and we calculated the accuracy and confusion matrix for each iteration.

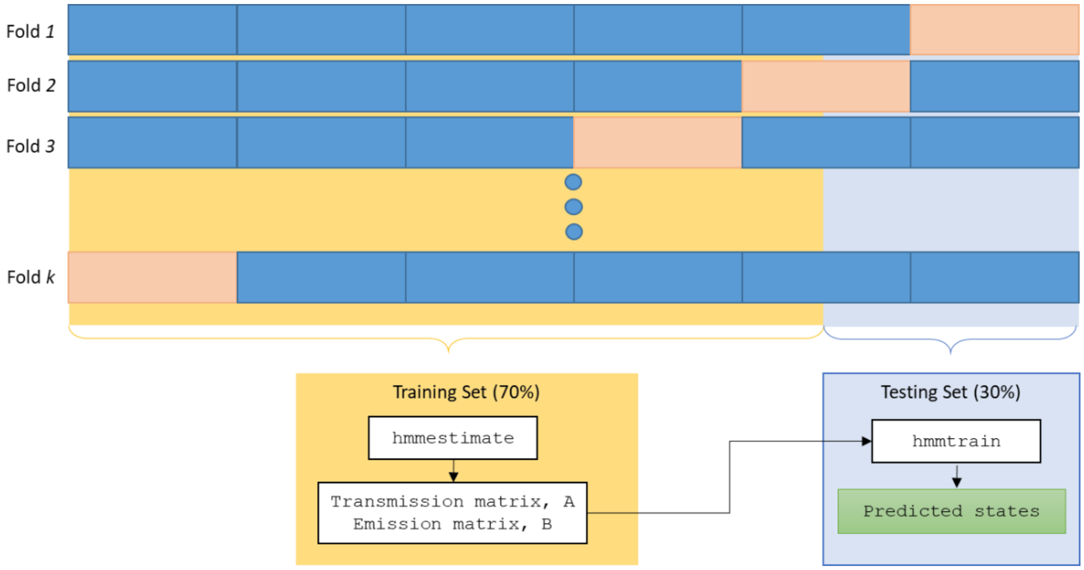


Figure 7-7: *k*-fold cross validation to train and test the HMM model

7.5.6 Evaluating Prediction Models Performance

The annotators labeled an additional two sessions using the same rules as stated in 7.5.2, to serve as ground truth to evaluate the rule-based prediction model and HMM. The newly labeled data were tested against the rule-based model and HMM offline and the predicted states from both models were compared to the hand-labeled data and presented in the next section.

7.6 Results and Discussion

7.6.1 HMM Training and Testing Results

We trained a HMM model with the data from the first four sessions that were hand-labeled with the participants' behavior using *k*-fold cross validation method to optimize the training output. We selected a HMM with the highest test accuracy as it may indicate that the model could perform reliably.

For HMM training, each iteration should include all possible observations. However, since our dataset were distributed based on real occurrences, we found that certain instances of the *k*-fold could not model a solution due to missing observations and we skipped that fold. We present in Table 7-4 the average accuracies and F-1 scores for *k* = 5, 6, 7, 8, 9, 10 based on the individual accuracy for the models that were successfully generated. The accuracy was calculated by comparing the states predicted by HMM to the hand-labeled states. We found that when *k* = 8, we achieved the highest average test accuracy of 95%. Next, we looked at the individual HMM within the *k* = 8 cross-validation results and selected the HMM with the highest overall results as shown in Table 7-5.

Table 7-4: Average accuracies of HMM models using k-fold cross-validation

***k*-value**	**Average Test Accuracy**	**Average F-1 Score**
5	91.6	92.6
6	92.7	93.5
7	93.3	94.2
8	**95.9**	93.7
9	94.3	95.0
10	88.6	90.6

Table 7-5: HMM models accuracies with *k* = 8

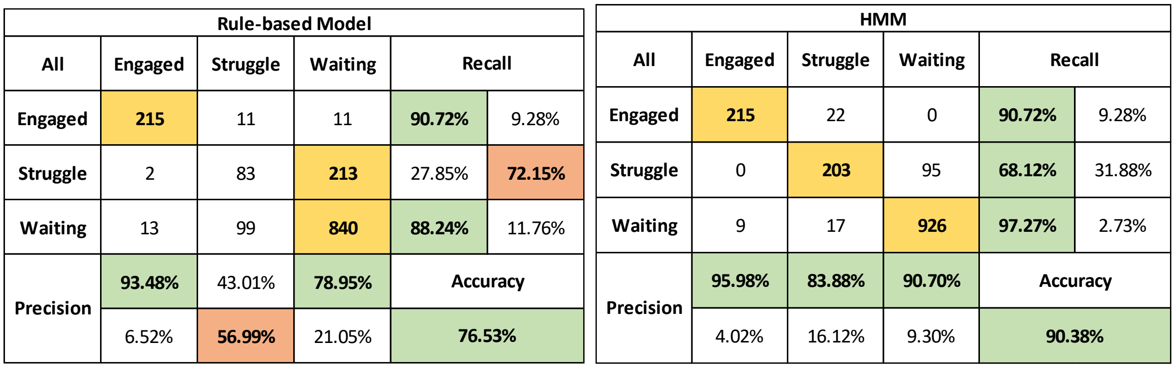
Fold (*k* = 8)	**Test Accuracy**	**F-1 Score**
1	80.53	83
2	97.69	98
3	98.02	98
4	**99.34**	**99**
5	n/a	n/a
6	88.45	90
7	n/a	n/a
8	n/a	n/a

7.6.2 Validation

The validation was done using data from the remaining two sessions that were hand-labeled after the four original sessions. For the HMM, we performed offline prediction using the model selected in the previous section (Model 4 from *k*-fold = 8). Table 7-6 compares the performance of rule-based prediction model and HMM prediction model and Figure 7-8 illustrates the confusion matrix of both prediction models.

Table 7-6: Validation results for rule-based prediction model and HMM

	Rule-based	**HMM**
Accuracy	76.53%	90.38%
Precision	71.81%	90.19%
Recall	68.93%	85.37%



Rule-based Model					
All	Engaged	Struggle	Waiting	Recall	
Engaged	215	11	11	90.72%	9.28%
Struggle	2	83	213	27.85%	72.15%
Waiting	13	99	840	88.24%	11.76%
Precision	93.48%	43.01%	78.95%	Accuracy	
	6.52%	56.99%	21.05%	76.53%	

(a)

HMM					
All	Engaged	Struggle	Waiting	Recall	
Engaged	215	22	0	90.72%	9.28%
Struggle	0	203	95	68.12%	31.88%
Waiting	9	17	926	97.27%	2.73%
Precision	95.98%	83.88%	90.70%	Accuracy	
	4.02%	16.12%	9.30%	90.38%	

(b)

Figure 7-8: Confusion matrix for (a) HMM and (b) Rule-based model

Overall, the HMM provided higher accuracy, precision, and recall of the participants' state compared to a rule-based model. When we look at the individual state as shown in Figure 7-8, both models performed really well for 'Engaged' state since the conditions for the '*Engaged'* state were quite simple and straightforward where both models could provide a reliable prediction. However, for '*Waiting'* and '*Struggling'* states, the rule-based model performed quite poor where the model predicted most of the '*Struggling'* state as '*Waiting'*. The inflexibility of rule-based model could have caused this. Rule-based model only allows one state for one set of condition, whereas real hand-labeled data would have instances where the same condition produced different outcomes based on the context of the task. For example, in one instance there was no gaze detected, no input device manipulated, and object was not moving, the rule-based would always predict the state as '*Struggling'*, but hand-labeled data could have labeled it as: i) '*Struggling'* - when the overall context at the time showed that the participant was indeed struggling, or ii) '*Waiting'* – when the participant was actually waiting for their partner to perform a task. If we keep rule-based model to predict participants' states, the feedback mechanism that the participants received would not be true to their actual state. A participant that is in '*Struggling'* state would not be prompted to seek assistance as the system would assume they are '*Waiting'* for their partner to complete a turn. On the other hand, the HMM prediction results for '*Waiting'* and '*Struggling'* states were good since the temporal information that was learned from the training was embedded within the state transition probability matrix, A, and emission probability distribution, B.

7.7 Conclusion and Future Work

Feedback mechanism in a virtual training environment is an important component to create a robust training environment that could improve learning outcomes. Recent research in this area explored incorporating human behavior together with task performance for improved feedback. Various methods have been explored to recognize or predict human behaviors such as machine learning algorithms, and probabilistic pattern recognition models. However, in a more complex training environment such as for teamwork skills training where dyads must collaborate and interact with each other in the same virtual space, the feedback mechanism would need to look at the changing dynamic of the interactions rather than a static approach. Motivated by this, our work presents the design of two models to predict human behaviors in a teamwork training environment. We wanted to compare the performance of a simple rule-based prediction model and a temporal-sensitive HMM against hand-labeled data in such a dynamic environment. Based on the validation results, we found that HMM had the best overall performance compared to the rule-based prediction model. Although the rule-based model had about the same accuracy as HMM for the '*Engaged'* state which shows that it is sufficient for a simple evaluation, for more complex states, '*Struggling'* and '*Waiting'*, it was difficult to scale up the rule-based model since it was quite rigid. HMM would fit better in a dynamic interaction as it offers more flexibility than a rule-based model. As we continue with our research in this area, future work will involve implementing this HMM in the collaborative tasks for real-time prediction of the participants' states.

Although the results are promising, it is important to highlight the limitations of the HMM design and important improvements in future works. First, the number of human states used to capture participants' collaboration behavior were basic. In future work, researchers would benefit by expanding the states, primarily for '*Waiting',* into more distinguished states to allow the researchers to better understand what is taking place in the collaboration. Second, the use of binary features was able to capture the simple states that were used. However, more complex analysis of the features would allow researchers to understand the collaboration better. For example, adding a dialogue act classification for the speech feature would better inform whether the participant said something because they needed help or sharing information, which can be represented by different states. Third, the number of labeled data used in the validation was relatively small. Future study with larger sample size, more states and features introduced would increase the understanding of collaborative interactions. Despite these limitations, results from the validation showed the advantage of HMM over rule-based prediction model in a dyadic collaborative interaction between autistic and neurotypical individuals.

References

[1] American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington, VA: American Psychiatric Association; 2013

[2] Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. Human factors, 50(3), 540-547.

[3] Amat, A. Z., Adiani, D., Tauseef, M., Breen, M., Hunt, S., Swanson, A., Weitlauf, A. S., & Sarkar, N. Design of a Virtual Reality-based Collaborative Activities Simulator (ViRCAS) to Support Teamwork in Workplace Settings for Autistic Adults. IEEE Transactions on Neural Systems and Rehabilitation Engineering [submitted-under revision]

[4] Green, C. S., & Bavelier, D. (2012). Learning, attentional control, and action video games. Current biology, 22(6), R197-R206.

[5] Zhang, L., Amat, A. Z., Zhao, H., Swanson, A., Weitlauf, A., Warren, Z., & Sarkar, N. (2020). Design of an intelligent agent to measure collaboration and verbal-communication skills of children with autism spectrum disorder in collaborative puzzle games. IEEE transactions on learning technologies, 14(3), 338-352.

[6] H. Zhao et al., "INC-Hg: An Intelligent Collaborative Haptic-Gripper Virtual Reality System," ACM Trans. Access. Comput. TACCESS, vol. 15, no. 1, pp. 1–23, 2022

[7] Fredriksson, T., Issa Mattos, D., Bosch, J., Olsson, H.: Data Labeling: An Empirical Investigation into Industrial Challenges and Mitigation Strategies, pp. 202–216 (11 2020). https://doi.org/10.1007/978-3-030-64148-1 13

[8] Kotov, A., Bennett, P. N., White, R. W., Dumais, S. T., & Teevan, J. (2011, July). Modeling and analysis of cross-session search tasks. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (pp. 5-14).

[9] Vondrick, C., Patterson, D., & Ramanan, D. (2013). Efficiently scaling up crowdsourced video annotation: A set of best practices for high quality, economical video labeling. International journal of computer vision, 101, 184-204.

[10] Hagedorn, J., Hailpern, J., & Karahalios, K. G. (2008, May). VCode and VData: illustrating a new framework for supporting the video annotation workflow. In Proceedings of the working conference on Advanced visual interfaces (pp. 317-321).

[11] Gaur, E., Saxena, V., & Singh, S. K. (2018, October). Video annotation tools: A Review. In 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) (pp. 911-914). IEEE.

[12] Vondrick, C., Patterson, D., & Ramanan, D. (2013). Efficiently scaling up crowdsourced video annotation: A set of best practices for high quality, economical video labeling. International journal of computer vision, 101, 184-204.

[13] Salah, A., Gevers, T., Sebe, N., Vinciarelli, A.: Challenges of human behavior understanding. vol. 6219, pp. 1–12 (12 2010). https://doi.org/10.1007/978-3-642- 14715-9 1

[14] Nikiforova, L. (2022). Model of Human Behavior Classification and Class Identification Method for a Real Person. Supplement.

[15] Sugimoto, M., Zin, T. T., Toriu, T., & Nakajima, S. (2011). Robust rule-based method for human activity recognition. International journal of computer science and network security, 11(4), 37-43.

[16] Sargano, A. B., Gu, X., Angelov, P., & Habib, Z. (2020). Human action recognition using deep rule-based classifier. Multimedia Tools and Applications, 79, 30653-30667.

[17] Grimm, M., Mower, E., Kroschel, K., & Narayanan, S. (2006, September). Combining categorical and primitives-based emotion recognition. In 2006 14th European Signal Processing Conference (pp. 1-5). IEEE.

[18] Storf, H., Becker, M., & Riedl, M. (2009, April). Rule-based activity recognition framework: Challenges, technique and learning. In 2009 3rd International Conference on Pervasive Computing Technologies for Healthcare (pp. 1-7). IEEE.

[19] Waltl, B., Bonczek, G., & Matthes, F. (2018). Rule-based information extraction: Advantages, limitations, and perspectives. Jusletter IT (02 2018).

[20] Abdelrahman, A. A., Strazdas, D., Khalifa, A., Hintz, J., Hempel, T., & Al-Hamadi, A. (2022). Multimodal Engagement Prediction in Multiperson Human–Robot Interaction. IEEE Access, 10, 61980-61991.

[21] Okada, S., Ohtake, Y., Nakano, Y. I., Hayashi, Y., Huang, H. H., Takase, Y., & Nitta, K. (2016, October). Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets. In Proceedings of the 18th ACM International Conference on Multimodal Interaction (pp. 169-176).

[22] H. Cuayáhuitl, S. Keizer, and O. Lemon, "Strategic dialogue management via deep reinforcement learning," 2015. [Online]. Available: arXiv:1511.08099. doi: 10.17861/6c6de69e-25ea-4836-b443-44b312354fac.

[23] Mihoub, A., Bailly, G., & Wolf, C. (2013, October). Social behavior modeling based on incremental discrete hidden Markov models. In International Workshop on Human Behavior Understanding (pp. 172-183). Springer, Cham.

[24] Zhang, D., Gatica-Perez, D., Bengio, S., & McCowan, I. (2006). Modeling individual and group actions in meetings with layered HMMs. IEEE Transactions on Multimedia, 8(3), 509-520.

[25] Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2020). Human emotion recognition: Review of sensors and methods. Sensors, 20(3), 592.

[26] A. Juliani et al., “Unity: A General Platform for Intelligent Agents.” arXiv, May 06, 2020. Accessed: Dec. 20, 2022. [Online]. Available: http://arxiv.org/abs/1809.02627

[27] D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. Journal of educational psychology, 105(4), 1082.

[28] D’Mello, S., Olney, A., Person, N.: Mining collaborative patterns in tutorial dialogues. Journal of Educational Data Mining 2 (01 2010)

[29] Schmidt, M., Laffey, J., Schmidt, C., Wang, X., Stichter, J.: Developing methods for understanding social behavior in a 3d virtual learning environment. Computers in Human Behavior 28, 405–413 (03 2012). https://doi.org/10.1016/j.chb.2011.10.011

[30] Basden, B., Basden, D., Bryner, S., Thomas, R.r.: Developing methods for understanding social behavior in a 3d virtual learning environment. J Exp Psychol Learn Mem Cogn 23(5), 1176–91 (09 1997). https://doi.org/10.1037//0278- 7393.23.5.1176

[31] Probabilistic predictive models offer flexibility and scalability compared to deterministic predictive models [CITE].

[32] Rabiner, L., & Juang, B. (1986). An introduction to hidden Markov models. ieee assp magazine, 3(1), 4-16.

[33] MathWorks, Inc. (1996). MATLAB: the language of technical computing: computation, visualization, programming

[34] MATLAB and Machine Learning Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States.

CHAPTER 8: CONTRIBUTIONS AND FUTURE WORK

Teamwork and EF are important life skills that can be developed from a young age to prepare individuals for future success in education, employment, and social life [1, 2, 3]. One way to effectively develop these skills is through training. Human-computer interaction (HCI) systems such as virtual reality (VR) has been widely studied by researchers for various applications from entertainment to training specific skills. Since VR-based systems can be engaging, many studies have explored the application of VR-based systems in supporting skills training for individuals with disabilities, including individuals with ASD. Among the specific skills that have been studied include literacy skills [4], social skills [5], and safety skills [6]. Although these studies have reported improvements in the specific skills they trained on, interaction in VR-based systems can be restrictive and not reflective of the complex real-world interactions, which may hinder the transfer of trained skills to the real world. Currently, there are no standardized measures for teamwork performance. Different organizations define and assess different criteria to represent teamwork performance [7, 8]. The majority of these assessments are based on qualitative evaluations performed by an observer or members of the team themselves, which are susceptible to bias and error. The use of quantitative measures of users' behavioral data is still understudied, specifically when multi-user interactions within HCI is involved. Motivated by these, my research work has contributed in many ways to address the issues mentioned above through the the design, development, and application of team-based activities in an intelligent collaborative virtual environment (CVE) that can encourage, assess, and efficiently support teamwork and executive functioning (EF) skills training, focusing on individuals with Autism Spectrum Disorder (ASD).

8.1 Technical Contributions

The first set of technical contributions is in the design and development of CVE systems to encourage collaboration in individuals with ASD. Existing method of training teamwork skills involves face-to-face interactions in a co-located environment. Setting up physical space to support real-world simulation for teamwork training scenarios can be resource-straining and costly [9]. A conventional VR-based system can simulate real-world training scenarios with minimal cost and effort. However, communication and interaction within VR-based systems are limited to one user interacting with the system. Additionally, for individuals with ASD, interacting face-to-face can be stressful and distract them from working on their skills [10]. In our CVE systems, we were able to design a variety of tasks with different scenarios across a wide range of virtual environments for teamwork training suitable for both children and adults with ASD. We integrated the use of a haptic device with a force gripper to add immersive experience for the participants. The CVE systems can be an alternative for individuals with ASD to practice and hone their communication and social skills, taking the loads off from therapists and clinicians. The CVE systems we built have four main contributions: 1) design and development of virtual collaborative games and team-based activities suitable for both children and adults with ASD, respectively, 2) integration of interactive multimodal sensors and devices to capture objective measures of the interaction, 3) design of a closed-loop feedback mechanism capable of providing simple acknowledgment and prompts based on real-time tracking of users' tasks performance, and 4) design and development of a haptic device with a force gripper as an interactive controller allowing users to 'feel' the virtual objects they are manipulating.

The second technical contribution is identifying dimensions of teamwork and executive functions suitable for collaborative task interaction and mapping of the dimensions to multimodal data to quantitatively measure teamwork and executive functions skills. Existing methods of evaluating teamwork and executive functions are done manually by a human observer or with limited automated tracking of activities through sensors and video recordings [8, 11, 12]. These observation methods are susceptible to bias and error. We conducted a literature review to identify the most relevant dimensions of teamwork and executive functions as there are no standardized measures available for teamwork and executive functions at this time. Quantitative measures from the multimodal data provided objective measures of the dimensions of teamwork and executive functions. The structured measures can complement observations made by human observers, minimizing bias and errors. Some of these quantitative measures are impossible for humans to measure through observation but can be easily acquired from sensors and system calculations, such as gaze duration and frequency, heart rate variability, the grip strength during manipulation, and frequency of button presses. Our contributions include: 1) identifying dimensions that can be used to represent teamwork and executive functions through literature surveys; and 2) mapping of multimodal measures to the dimensions identified for teamwork and executive functions.

The third set of technical contributions is in the design and development of an intelligent agent within the CVE systems. There are three main objectives of the intelligent agent: 1) as a collaboration partner when a human partner is not present; 2) to monitor and evaluate users' teamwork behavior; and 3) to provide individualized real-time feedback to users based on individual and combined performance. In the CVE collaborative puzzle games, we designed a conversational agent that can actively participate in the collaboration, where the agent can initiate conversation and also respond accordingly to the users. Using a rule-based design in a simple HCI system is sufficient to monitor, evaluate, and respond to users. However, in order to support complex interaction during collaboration, a rule-based design was not scalable and as robust to the new rules introduced in multi-user interactions. As such, we designed a more robust and flexible method to interpret and predict human behavior using a Hidden Markov Model (HMM). Our intelligent agent design for the CVE systems has the following contributions: 1) a dialogue manager model capable of processing user's speech, detecting user's intention, and generating speech; 2) a probability model trained with ground truth data to predict participants' behavior in collaborative interactions; and 3) design and develop a finite state machine to provide real-time individualized feedback to each user based on their predicted state.

8.2 Societal Contributions

Other than the technical contributions, this work also has societal contributions. The work have been validated through feasibility and pilot studies with human participants. This work: 1) provided autistic individuals with access to a teamwork training simulator, 2) raise awareness in non-autistic individuals of the collaborative patterns and behaviors of autistic individuals, 3) support human observers in assessing teamwork skills through the use of the CVE system, 4) allow researchers to identify relevant dimension that can be used to assess teamwork and executive functions in an employment setting, and 5) expand the application of multimodal data analytics for teamwork and executive functions evaluation. As for contribution to research related to autism, results from the studies can be used to understand better how HCI systems effect skills learning in adults with ASD and ways we can improve the systems to support them better.

8.3 Future Work

Research in the area of teamwork skills training and computer-based collaborative interactions for autistic individuals offers many opportunities for new discoveries. This research has paved the way to understand and explore the opportunity teamwork training can offer.

One future direction that we would like to focus on is the HMM prediction model and the feedback mechanism. In our pilot study, we introduced three behavioral states which were “Engaged”, “Struggle”, and “Waiting” to drive the feedback mechanism in the collaborative interaction. However, in some instances, it would be better to have more detailed information of this behavior. For instance, understanding whether the participant is struggling with the task or struggling with communicating with their partner would allow the intelligent agent to provide a more targeted prompt. So far, the observations from the multimodal data only provide task-related struggle information from the controller information. Additional data processing could introduce a new observation that tracks whether the participant has been initiating conversation or responding to their partner’s initiation. If they are not initiating or responding, the agent could prompt them to do so.

Another area of future development is automating the assessment and evaluation of teamwork and executive functions. Currently, we are performing multimodal analytics for the dimensions of teamwork and executive functions offline, and it involved laborious manual work. Researchers can train machine learning models to analyze and rate the skills based on existing data we have from studies we conducted.

In our study comparing the performance of a control group and a training group, we were able to observe immediate potential of the training tasks. Although we observed a positive improvement in both task performance and communication skills for the participants in the training group, there were also improvements observed in the control group. Future research may benefit from a longitudinal study that analyzes the effect of longer training sessions with the collaborative training tasks.

References

[1] Brown, T. E. (2013). A new understanding of ADHD in children and adults executive function impairments. New York, NY: Routledge.

[2] Trilling, B., & Fadel, C. (2009). 21st century skills: Learning for life in our times. John Wiley & Sons.

[3] Salas, E., Fowlkes, J. E., Stout, R. J., Milanovich, D. M., & Prince, C. (1999). Does CRM training improve teamwork skills in the cockpit: Two evaluation studies. Human Factors, 41(2), 326-343.

[4] Colby, K. M. (1973). The rationale for computer-based treatment of language difficulties in nonspeaking autistic children. Journal of autism and childhood schizophrenia, 3(3), 254-260.

[5] Bernardini, S., Porayska-Pomsta, K., & Smith, T. J. (2014). ECHOES: An intelligent serious game for fostering social communication in children with autism. Information Sciences, 264, 41-60.

[6] Saiano, M., Pellegrino, L., Casadio, M., Summa, S., Garbarino, E., Rossi, V., ... & Sanguineti, V. (2015). Natural interfaces and virtual environments for the acquisition of street crossing and path following skills in adults with Autism Spectrum Disorders: a feasibility study. Journal of neuroengineering and rehabilitation, 12(1), 1-13.

[7] Sung, C., Connor, A., Chen, J., Lin, C. C., Kuo, H. J., & Chun, J. (2019). Development, feasibility, and preliminary efficacy of an employment-related social skills intervention for young adults with high-functioning autism. Autism, 23(6), 1542-1553.

[8] Lai, E., DiCerbo, K., & Foltz, P. (2017). Skills for Today: What We Know about Teaching and Assessing Collaboration. Pearson.

[9] Grob, C. M., Lerman, D. C., Langlinais, C. A., & Villante, N. K. (2019). Assessing and teaching job-related social skills to adults with autism spectrum disorder. Journal of applied behavior analysis, 52(1), 150-172.

[10] Parsons, S., Beardon, L., Neale, H. R., Reynard, G., Eastgate, R., Wilson, J. R., ... & Hopkins, E. (2000, September). Development of social skills amongst adults with Asperger’s Syndrome using virtual environments: the ‘AS Interactive’project. In Proc. The 3rd International Conference on Disability, Virtual Reality and Associated Technologies, ICDVRAT (pp. 23-25).

[11] Echeverria, V., Martinez-Maldonado, R., & Buckingham Shum, S. (2019, May). Towards collaboration translucence: Giving meaning to multimodal group data. In Proceedings of the 2019 chi conference on human factors in computing systems (pp. 1-16).

[12] Bozgeyikli, L., Bozgeyikli, E., Raij, A., Alqasemi, R., Katkoori, S., & Dubey, R. (2017). Vocational rehabilitation of individuals with autism spectrum disorder with virtual reality. ACM Transactions on Accessible Computing (TACCESS), 10(2), 1-25.