Design and Applications of Intelligent Human-Computer Interaction Systems for Autism Spectrum Disorder intervention

By

Lian Zhang

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

May 31, 2018

Nashville, Tennessee

Approved:

Nilanjan Sarkar, Ph.D.

Zackary E. Warren, Ph.D.

Gabor Karsai, Ph.D.

Douglas Fisher, Ph.D.

D. Mitchell Wilkes, Ph.D.

Amy S. Weitlauf, Ph.D.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere appreciation to my advisor, Dr. Nilanjan Sarkar, for his invaluable advices and help during my Ph.D. study. I would also like to thank Dr. Zachary Warren and Dr. Amy Weitlauf for their help in conducting the meaningful interdisciplinary research. Furthermore, I would like to thank other members of my dissertation committee: Dr. Gabor Karsai, Dr. Douglas Fisher, and Dr. Mitch Wilkes, for their insight and feedback that improved this dissertation work significantly.

I would like to thank Ms. Amy Swanson at the Vanderbilt Kennedy Center Treatment and Research Institute of Autism Disorder (TRIAD) for her help on user study design and participant recruitment. I want to acknowledge the excellent colleagues, Joshua Wade, Jing Fan, Dayi Bian, Huan Zhao, Ashwaq Zaini Binti Amat, Guangtao Nie, Zhaobo Zheng, and many others from the Robotics and Autonomous Systems Lab who provided dedicated collaboration and resourceful help in this work. Also, I would like to thank Nicole Bardett and Anna Pasternak at the TRIAD for their professional help in the user studies.

I would like to express my deep thank to my father, my mother, my sister, my brother, and my significant other -- Qiang Fu. They have been by my side throughout my Ph.D. study. I could not have accomplished this goal without their support and encouragement.

Acknowledge the funding support by the Hobbs Society Grant from the Vanderbilt Kennedy Center, the National Science Foundation under Grant 0967170, and the National Institute of Health under Grant 1R01MH091102-01A1, and 1R21MH111548-01.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER I. INTRODUCTION

1.1. Overall Goal of this Work

Autism Spectrum Disorder (ASD) is a group of neurodevelopmental disabilities characterized by pervasive impairments in social interaction, communication, and atypical patterns of behaviors (Association 2000). The estimated prevalence of ASD is 1 in 68 in the United States (Developmental and Investigators 2014), and the individual incremental lifetime cost associated with ASD is over $3.2 million (Peacock et al. 2012). Although there is no single accepted intervention, treatment, or known cure for ASD, cumulative literature suggests behavioral and educational intervention programs have the potential to positively impact the lives of children with ASD and their families (Rogers 1998; Cohen et al. 2006). However, the lack of access to expert clinicians and the huge associated cost of traditional intervention are considered limitations of existing intervention programs. Therefore, the development of inexpensive and effective assistive therapeutic tools for ASD intervention is urgently needed.

Given recent technological advances, it has been argued that specific computer applications could be harnessed to provide low-cost and novel clinical treatments for children with ASD (Moore et al. 2005). Human Computer Interaction (HCI) is a field that studies interactions and communication between a human and a computer. HCI applications have been shown to be engaging to children with ASD (Pennington 2010). As such, researchers have developed a wide range of HCI applications in order to investigate social and communication behaviors of the children (Ramdoss et al. 2012). The strengths of these HCI applications include controllability, repeatability, and safety (Parsons and Cobb 2011; Burke et al. 2010). Specifically, they can provide controlled and safe environments where specific social and communication behaviors can be tested and taught repeatedly. However, most of the existing HCI systems for ASD intervention are in a single-user mode and focus on interactions and communication between a user and a computer. Such systems are often limited by the programming burden of realizing flexible social communication paradigms (M Schmidt et al. 2011).

A Collaborative Virtual Environment (CVE), which is a computer-based, distributed, virtual space for multi-user to interact with one another (Benford et al. 2001), is a sub-field of HCI that allows interactions between multiple humans via computers (Baecker 2014). It offers an effective way to facilitate flexible communication between real users within a controllable system (Leman 2015; Reynolds et al. 2011). Such systems offer promising platforms for individuals with ASD to practice their social skills in controlled environments with realistic settings. Previous literature has reported positive impacts of CVE systems on social skills of individuals with ASD (Ben-Sasson et al. 2013; Hourcade et al. 2012). However, existing CVE systems for ASD intervention had limitations to encourage collaboration between real-users, which is important aspects of social communication of the population. Therefore, one of the important goals of this research is to design CVE systems to promote collaboration between children with ASD and their peers. Another limitation of existing HCI systems for ASD intervention is related to the challenge of evaluating behaviors of individuals with ASD in HCI systems, particularly given the fact that conventional methods, manual video/audio coding, of doing so are time-intensive and laborious. In order to overcome this issue, we explored a novel way to perform automatic evaluation of interaction in this work.

One type of technology with the potential for automatic evaluation is an intelligent HCI system, which can perceive users' behaviors using artificial intelligent methodologies (Xu and Wang 2006). One of the distinctive features of an intelligent HCI system is user modeling (Brusilovsky and Millán 2007). The goal of user modeling is to understand specific behaviors of a user. A few intelligent HCI systems have been developed to understand behaviors in the ASD population, such as their affective and cognitive states (Bian et al. 2016), eye gaze pattern (Wade et al. 2016), and communication skills (Bernard-Opitz et al. 1999). These systems not only provide low-cost, accurate, and meaningful measurements of social behaviors; they also feed measurements back into the systems in order to facilitate continued engagement and enhance learning. Despite these potential benefits, intelligent HCI systems applied to automatic evaluation of interaction are not yet matured and require further development. Therefore, the primary goal

of my research is to design and apply intelligent HCI systems for ASD intervention in order to fill the gap in the existing literature.

My research focuses on the design and application of HCI systems, especially CVE systems and intelligent systems, for ASD intervention. The goals of this work are to: i) design CVE systems in order to encourage collaborations between real-users, ii) design intelligent HCI systems in order to automatically measure interactions in the CVE systems, and iii) explore intelligent HCI systems to measure cognitive load of the population. The main contributions of this work include a novel platform for children with ASD to practice collaborative interactions and communication with their TD peers, as well as an intelligent system to automatically measure both communication and collaboration skills of the children during the interactions. In addition, this work contributes to this research area by providing a framework to automatically measure cognitive load of the targeted population. In the following sections of this chapter, I first present a detailed survey of existing literature on HCI systems, including both single-user and multi-user systems, for ASD intervention in Section 1.2. Section 1.3 reviews studies on intelligent systems to measure outputs of HCI intervention systems. Section 1.4 summarizes my research.

1.2. Literature review on HCI systems for ASD intervention

There are several reasons why HCI systems may be particularly effective for ASD intervention. The primary reason is due to their controllability, repeatability, and safety (Parsons and Cobb 2011; Burke et al. 2010). Individuals with ASD often have differences in sensory perception, which may lead to difficulties in screening out unnecessary information in complex environments (Williams et al. 2002). HCI systems can filter out secondary information and present only primary information to the individuals for a targeted intervention. As a result, these controllable systems could be particularly suitable for ASD intervention. In addition, many individuals with ASD have a natural affinity for controlled environments provided by computers (Moore et al. 2005), and exhibit a high level of engagement in these environments (Lahiri et al. 2015). HCI systems that can engage them may thereby enhance learning. Because of these

reasons, HCI technology appears well-suited for creating interactive skill training paradigms in core areas of impairment for children with ASD.

Because the individuals show deficits in social interactions and communication, the HCI systems used for ASD intervention try to mitigate these deficits (Moore and Taylor 2000). With this particular purpose, these systems usually are required to i) offer targeted stimuli in order to elicit users' specific behaviors, or ii) provide proper feedback in order to enhance learning. Although HCI systems have been widely used in our daily life (e.g., games, virtual environments, and online services), these cannot be directly used for ASD intervention because they are not adaptable for targeted controllable interactions and are unable to provide appropriate feedback (Caltagirone et al. 2002; Livingstone et al. 2008).

Based on their targeted interactions, existing HCI systems for ASD intervention can be classified into three categories. The first category includes HCI systems that support multimedia-based interactions. Previous intervention systems have used multimedia, i.e., image, audio, and video, in order for individuals with ASD to practice social interactions and communication (Colby 1973). Compared to the traditional therapist-based intervention, these HCI systems can allow social interactions and communication to be practiced individually and repeatedly (Wainer and Ingersoll 2011). However, these HCI systems usually can only support one-way interaction, i.e., users understanding multimedia contents displayed by computers. In addition, these HCI systems often utilized two-dimensional images for the interaction, which are less realistic compared to three-dimensional objects in real life (Council 1994). The second category includes Virtual Reality (VR)-based systems that support interactions between users and virtual environments/avatars using three-dimensional virtual objects. VR refers to the computer-generated simulation of a world and/or engaging users in the simulated world. Compared to multimedia-based HCI systems, VR-based systems can support two-way interactions, i.e., both a user acting in a virtual environment and the environment responding to the user actions. The simulated environments and the two-way interactions of VR-based systems may lead to a higher level of engagement. However, realizing completely realistic interactions with virtual environment is all but impossible with current computer-assisted technologies (Council 1994), and simulating flexible and unrestricted social communication

using programmed virtual avatars is challenging from a technical point of view (Suchman 1987). Therefore, the less than realistic interactions and insufficiently flexible communication are limitations of this kind of VR-based system. The third category includes multi-user HCI systems that support interactions among multiple real persons within virtual environments. A multi-user HCI system has the ability to facilitate realistic interactions and flexible communication among real users.

1.2.1. Multimedia-based intervention systems

Several studies have shown that multimedia-based HCI systems can be useful for ASD intervention. Colby (1973) designed a computer program that provided auditory and visual feedback in response to pressing a letter on a keyboard (Colby 1973). The computer program was designed in order to encourage nonverbal children with ASD to speak. Thirteen out of 17 participants showed linguistic improvements after using the computer program. This pioneering work indicated the potential usability of multimedia-based HCI systems for ASD intervention. Other researchers have also developed multimedia-based HCI systems to engage individuals with ASD in their reading and writing (Williams et al. 2002), sentence construction (Yamamoto and Miya 1999), and functional conversation (Hetzroni and Tannous 2004). The multimedia-based HCI systems usually were designed to offer multimedia contents, such as image, audio, and video, for remediation of ASD's deficits (Bellini and Akullian 2007). The capability of these systems to present attractive and engaging audio/video content as feedback serve as their advantages compared to traditional teaching and training methods.

Multimedia-based HCI systems have also been developed to study the effects of different kinds of multimedia contents on core deficits of ASD. For example, Golan and colleagues designed software to promote emotion recognition of children with ASD (Golan and Baron-Cohen 2006). The software displayed faces with different emotions using different kinds of films, i.e., films showing eye area only, films with and without voice, and films with contextual information. A total of 54 children with ASD and 24 TD children participated in this study for several months. Statistical analyses showed that children who utilized the software improved significantly more in emotion recognition than children who did not

use the software. The study also demonstrated the importance of key characteristics, such as voices, eyes, and social situation, in emotion recognition within the system.

Other studies in this area have also investigated effects of different multimedia contents under different conditions. Ploog and colleagues designed a multimedia-based HCI system to investigate the effects of different linguistic components and prosody on attracting attention of children with ASD (Ploog et al. 2009). Bernard-Opitz and colleagues investigated effects of systems with and without image feedback on individuals with ASD when solving social problems (Bernard-Opitz et al. 2001). Heimann and colleagues designed a system that could display different types of content, including animation, video, and voice, to improve reading and communication skills of children with ASD (Heimann et al. 1995). These studies indicated the importance of specific multimedia-contents in learning certain skills. Although these early multimedia-based HCI systems have shown promising results in ASD intervention, the majority of them can only support simple interactions, i.e., facial expression recognition and vocabulary learning. As a result, this kind of HCI system have limitations in the effects of the intervention systems.

1.2.2. Virtual reality-based intervention systems

Virtual Reality (VR) based systems for ASD intervention began in 1990s (Parsons and Cobb 2011). VR refers to using computer-technology to generate a virtual world into which users can be immersed (Rheingold 1991). The virtual world is usually responsive to user's actions. Various displays, including immersive head mounted displays (HMD), were employed in the early phases of VR-based systems for ASD intervention. However, HMD often were rated as heavy and caused discomfort in these studies (Parsons et al. 2004). As a result, desktop-based VR were preferred to HMD-based VR when used for ASD intervention (Wang and Reid 2010). Virtual avatars are programmed virtual characters in the virtual world and have been used to understand and enhance social and communication skills of individuals with ASD. Moore and colleagues designed a virtual avatar that can have four different facial expressions, i.e., happy, sad, angry and frightened. The virtual avatar was developed in order to evaluate the ability of children with ASD in identifying emotions and making inference on emotions (Moore et al. 2005).

Thirty-four individuals with ASD participated in their study and over 90% of the participants accurately recognized the emotions portrayed by the avatar. This study, as one of the early studies in this area, indicated the usability of virtual avatars in the emotion recognition of the targeted population. Virtual avatars have also been used to investigate emotion recognition skills (Esubalew Bekele et al. 2014), mental state recognition skills (Konstantinidis et al. 2009), and eye gaze behaviors (Mineo et al. 2009) of the targeted population. These virtual avatars, which had specific social communication functionalities, such as the capabilities to speak and make facial expressions (Esubalew Bekele et al. 2013), were designed for social communication skills training. VR-based systems that focused on simulating daily-life scenarios have also been applied for individuals with ASD to practice their daily-life skills (Parsons et al. 2004; Esubalew Bekele et al. 2014).

Virtual environments can, to some extent, replicate real social worlds for individuals with ASD to practice specific behaviors within the environments (Moore et al. 2005). For example, Parsons and colleagues designed a virtual café to understand social appropriateness of children with ASD (Parsons et al. 2004). The virtual objects in the environment could prompt interactions by providing image- and verbal-responses when a child approaches to them. Twelve children with ASD as the experimental group and 12 TD children as the control group were involved in their study. These participants explored the virtual café by navigating the environment and interacting with the virtual objects. It was found that some of the children with ASD exhibited inappropriate behaviors in the virtual environment, e.g., they were more likely to bump into or walk between other people, as compared to their TD peers. Despite these differences, these children with ASD could easily use the virtual environment.

Virtual environments have two advantages relative to real world settings in that they are safe and controllable. Because of the safety, virtual environments have been used for individuals with ASD to learn some risky daily life skills, such as street-crossing (Josman et al. 2008) and driving (Wade et al. 2016). The controllability aspect of virtual environments, on the other hand, enables individuals with ASD to practice specific social skills, such as shopping (Lányi and Tilinger 2004), with varying complexity. Despite these advantages, this kind of system, which involves a single user within a virtual

environment, usually can only support preprogrammed interactions and communication between a user and the environment (M Schmidt et al. 2011). This kind of interaction is often unrealistic and inflexible. This is considered as one of limitations in VR-based HCI systems for ASD invention.

1.2.3. Multi-user intervention systems

Multi-user HCI systems can support collaborative interactions and natural communication between real users within a controlled environment. The ability to support realistic interactions and flexible communication between real users is an advantage of multi-user HCI system for ASD intervention. Multi-user HCI systems can be categorized into co-located systems and geographically distributed systems. The majority of multi-user HCI systems for ASD intervention were co-located systems, which utilized multi-touch devices in the same location to investigate collaborative behaviors in the ASD population. Gal and colleagues developed a multi-touch device, named StoryTable, to evaluate collaborative interactions between children with ASD (Gal et al. 2009). The device held a variety of backgrounds and settings, and children with ASD could collaboratively select these backgrounds and settings to form a story. Eight children with ASD were involved in the study, and their collaboration levels were evaluated using a friendship observation scale (Bauminger et al. 2005). It was found that these participants had an increased frequency of complex play after using the device. Using co-located multi-touch devices, previous literature also investigated other collaborative behaviors of children with ASD, such as sharing (Curtis and Lawson 2001), turn taking (Zancanaro et al. 2007), and collaborative play (Ben-Sasson et al. 2013). These studies demonstrated the usability of these co-located multi-touch devices in understanding collaborative interactions of individuals with ASD.

The other category of multi-user HCI systems for ASD intervention includes geographically distributed systems, named Collaborative Virtual Environments (CVEs). These systems can support collaborative interactions between users in different locations. One obvious advantage of the distributed design is that it can offer more chances for individuals with ASD to interact with people in different locations (Millen et al. 2012). Additionally, co-located multi-user HCI systems required all users to be in

the same location for face-to-face interactions. Face-to-face interactions may be initially difficult for children with ASD due to differences in social understanding and skills. Thus, a distributed CVE, which allows users to interact in different locations, provides a platform for children with ASD to comfortably interact with their peers (Millen et al. 2012). Previous research has investigated using CVEs to promote social communication skills of individuals with ASD. Schmidt and colleagues developed a 3D CVE system, named iSocial, to investigate the individuals' skills of reading facial expressions and predicting other's thoughts (Stichter et al. 2014). Results of the study, which involved 11 children with ASD, demonstrated that a social competence curriculum could be delivered in the virtual environment. CVE systems have also been applied to investigate other social behaviors, such as social competence (M Schmidt et al. 2011), face-to-face communication (Millen et al. 2012), and empathy (Cheng et al. 2010). However, collaborations between real-users, which are important aspects of social communication in the population, have seldom been investigated in the literature. Therefore, one of the main goals of this work is to design CVE systems that can encourage collaborative interactions between individuals with ASD and their TD peers.

1.2.4. Limitations of existing HCI systems for ASD intervention

As described above, existing HCI systems can provide safe and controllable environments for users to practice their social interactions and communication skills. Furthermore, CVE systems provide a promising platform to investigate collaborative interaction and flexible communication between multiple users across the internet. However, the use of these HCI systems for pragmatic intervention has been limited by one fundamental challenge. In particular, most of them evaluated users' behaviors using a human-coding methodology, which is laborious and time-intensive. Autonomous systems that can provide automatic, consistent, and unbiased measurements of meaningful aspects of social interactions and communication within HCI systems will reduce the substantial time- and cost-effect associated with the human coding methodology. One way to develop such an autonomous system is to apply sensor technology to collect data of individuals with ASD and then apply artificial intelligence methodology to

understand their behaviors from the collected data. This type of intelligent system could address the limitations of conventional HCI systems in evaluating social behaviors of individuals with ASD. The next section discusses existing intelligent HCI systems for ASD intervention.

1.3. Next Generation of HCI Systems for ASD Intervention-Intelligent HCI Systems

An intelligent system is a computing system that can automatically perceive users/environments and dynamically interact with the users/environments (Xu and Wang 2006). Intelligent HCI systems with the capability to automatically detect behaviors of the users are meaningful for ASD intervention for several reasons. First, intelligent systems can automatically evaluate the behaviors and, therefore, reduce labor, time, and cost associated with a traditional human-coding methodology. It also reduces the risk of personal biases associated with the human-coding methodology. Second, some individuals with ASD have difficulties in understanding their own emotional and mental states (Rajendran and Mitchell 2007). An intelligent system that can automatically measure affective and cognitive states of individuals with ASD based on their physiological or gaze measurements, rather than self-report, has the potential to improve the measurement accuracy (Ozonoff and Strayer 2001). Finally, it is hard for human coders to accurately evaluate some implicit interaction cues, such as eye gaze. Intelligent systems may have the potential to measure these implicit interaction cues using accurate sensors.

Existing intelligent HCI systems usually understand individuals' behaviors by analyzing corresponding sensor signals (Brusilovsky and Millán 2007). A few studies have been conducted by researchers in this area to measure specific behaviors of individuals with ASD from their physiological (Bian et al. 2016), eye gaze (Wade et al. 2016), and speech (Bernard-Opitz et al. 1999) signals. These studies differed in terms of used sensors, targeted user behaviors, number of participants, and research goals. However, they represent a primary push for future research trends in HCI systems for ASD intervention. These systems can be classified into three categories based on the used sensor signals: physiological data-based, eye gaze data-based, and audio data-based. In what follows, we review works in

each of these categories by i) listing several important works in the category, and ii) describing in detail a study with a large sample size from each category.

1.3.1. Physiological data-based intelligent HCI systems

There is a considerable amount of work that recognizes affective and cognitive state of individuals with ASD from their physiological signals, such as heart rate, body temperature, and electromyogram activity (Bian et al. 2015; Lahiri et al. 2015; Nuske et al. 2014). Affective and cognitive state recognition is a core component of effective educational programs, which have the potential to engage users and reduce their frustration (Novak et al. 2012). However, some individuals with ASD have difficulties in correctly recognizing and reporting their own affective and cognitive states. Therefore, researchers have applied artificial intelligence methodologies to automatically recognize their affective and cognitive states from physiological signals. Liu and colleagues designed computer-based cognitive tasks (i.e., an anagram solving task and a pong playing task) to trigger users' different affective states, including enjoyment, anxiety, and engagement (Liu et al. 2008). They measured the users' affective states from multiple physiological signals, such as cardiovascular activity, electro-dermal activity, electromyogram activity, and peripheral temperature. These signals were recorded from the users when they participated in the cognitive tasks. Six young children with ASD were involved in their study. Results of the study indicated that these physiological signals could be used to measure the participants' affective states. Other works in this category have also utilized physiology signals for affective and cognitive state recognition under different conditions. For example, Bian and colleagues utilized physiological signals to measure affective states when individuals with ASD drove with a driving simulator (Bian et al. 2015). Kuriakose and colleagues measured anxiety levels of children with ASD in a VR-based social communication system (Kuriakose and Lahiri 2015).

1.3.2. Eye gaze data-based intelligent HCI systems

Eye gaze information of the individuals has been analyzed in order to understand their eye contact, visual attention, engagement, and eye gaze patterns. Eye contact, which may be absent or overly intense

in the individuals, is a form of non-verbal communication and has a large influence on their social behaviors (Stone et al. 1997); while visual attention guarantees safety in daily activities of the individuals (Lee et al. 2007). Therefore, researchers have analyzed eye gaze signals of the individuals to understand their eye contact patterns (Escobedo et al. 2012; Al-Omar et al. 2013) and visual attention (E Bekele et al. 2016). Atypical eye gaze patterns of the individuals have been demonstrated by multiple studies. Neumann and colleagues have designed a computer program to investigate eye gaze patterns of the targeted population when they recognized facial expressions (Neumann et al. 2006). The computer program could display blurred images of human faces with different kinds of facial expressions, i.e., fear and happiness. An eye tracker was applied to track users' eye gaze information when they used the program. Ten individuals with ASD and ten TD individuals were involved in their study. They found atypical gaze patterns in the individuals with ASD compared to the TD participants. In particular, these individuals with ASD fixated on the location of the mouth more than those TD individuals. In addition, Rutherford and colleagues have also found atypical eye-gaze patterns when children with ASD looked at images showing different emotions (Rutherford and Towns 2008). Reimer and colleagues reported atypical gaze pattern of children with ASD when using a driving simulator (Reimer et al. 2013). Eye gaze information, such as blink rate and pupil diameter, can reflect changes in engagement and emotion (Woolf et al. 2009). In this context, a study by Lahiri and colleagues have analyzed eye gaze signals of children with ASD to evaluate their affective states in a VR-based social communication system (Lahiri et al. 2015; Reimer et al. 2013). In our work, eye gaze data were combined together with other physiological data to recognize affective state of the ASD population.

1.3.3. Audio data-based intelligent systems

The category of "audio data-based intelligent systems" includes systems that can automatically analyze collected audio data of users (i.e., spoken language) in order to understand their communication skills. Individuals with ASD often have verbal communication difficulties. Intelligent systems that can understand their communication skills by analyzing their audio data, and/or enhance their specific aspects

of communication skills based on the understanding, have potential therapeutic benefits. However, automatically cataloguing, understanding, and responding to verbal communication skills is challenging, given the fact that designing a computer program that can understand and simulate unrestricted naturalistic conversation of humans (i.e., the Turing test) is problematic from a technical point of view. As such, existing intelligent HCI systems in this area have investigated only simple language components, such as vocabulary and sentence construction, in the ASD population. Bernard and colleagues developed multiple computer-assisted interactive tasks in order to encourage non-verbal children with ASD to speak (Bernard-Opitz et al. 1999). Ten non-verbal children with ASD were involved in the study. They were required to pronounce specific words within the tasks. The system could recognize their pronunciation using a speech engine (IBM speech viewer) and then provide graphical feedbacks based on their pronunciation. Results indicated that participants in the computer-assisted session had significantly greater vocal imitation as compared to participants in personal instruction session. Islam and colleagues developed an intelligent HCI system to improve vocabulary of children with ASD. Their system could recognize simple words, which were used by a user to name a object, and then offer feedback based on the recognized words (Islam et al. 2013). Ketterl and colleagues designed a similar system aimed at improving speech intelligibility of children with ASD. The speech intelligibility refered to the proportion of a speaker's speech that a listener can readily understand (Ketterl et al. 2011). Anwar and colleagues have designed a system that could display multiple images simultaneously in order for a user to construct a sentence to describe these images (Anwar et al. 2011). In their system, a speech enginee was used to convert a user's speech to text, and a human instructor was involved to provide feedback based on the text. These intelligent HCI systems focused on understanding simple language components of individuals with ASD from collected audio data and have shown positive impacts on their communication skills. However, intelligent HCI systems that can understand communication patterns of user-user interactions have not yet been developed to the best of our knowledge.

1.3.4. Limitations of existing intelligent HCI systems for ASD intervention

Existing intelligent HCI systems for ASD intervention applied artificial intelligence methodologies in computer-based interactive systems in order to interact with individuals with ASD, evaluate their behaviors and affective states, and adapt the systems' key components to enhance the individuals' learning efficiency. Studies in this area have analyzed multiple sensor signals, such as physiological, eye gaze, and audio signals, to measure their behaviors and affective states. In particular, physiological signals were used for affective state recognition. Eye gaze signals were analyzed to understand their atypical gaze patterns and visual processing. Audio signals have been investigated to understand specific aspects of communication skills of the individuals. Even though these systems have proved the usability of intelligent HCI systems in measuring specific social behaviors, two major limitations exist in the current intelligent HCI systems for ASD intervention:

1) The open-ended CVE systems pose no restriction in verbal communication between real users. As such, subsequent manual coding of interactions is necessary to understand patterns of communication for meaningful measurement and intervention. This creates a resource burden and limits realistic scale-up of the paradigm. An intelligent system that can automatically yield quantitative metrics of social communication and collaborative interactions within the systems may provide a way to address this challenge. Unfortunately, such systems that can automatically evaluate collaboration and communication skills of individuals with ASD in CVE systems have not been studied yet, to the best of our knowledge.
2) Although intelligent HCI systems have been investigated to recognize affective and cognitive states, most of these systems utilized single-modality signals, i.e., signals from single independent channel of sensor, to recognize the affective and cognitive states. Combining multimodal information may lead to higher accuracy in understanding the states. However, multimodal information fusion technologies have seldom been investigated for affective and cognitive state recognition in the literature.

This research provides a novel contribution by addressing these gaps in existing work related to the use of intelligent HCI systems for ASD intervention.

1.4. Overview of the Dissertation Research

My research focuses on the design and application of HCI systems, especially CVEs and intelligent systems, for ASD intervention. Most of existing HCI systems in this area were designed for individuals with ASD to practice their skills by interacting with computer programs. However, the interactions between users and computer programs were usually limited, with weak transfer of the skills learned within systems to real world settings. Additionally, these systems often utilized a human-coding method to measure the users' within-system behaviors. This human-coding method requires significant time, costs and efforts, limits the precision of the measurements, and restricts system capability for real-time feedback. In order to address these limitations, we designed and applied HCI systems for both intervention and measurements. In particular, our HCI systems had the potential to provide intervene on individuals' behavior and communication skills, as well as proposed efficient within-system measures to index the behaviors. In what follows, we briefly introduce each of our research studies.

1.4.1. Collaborative virtual environment systems

Traditional HCI systems that support interactions between a user and a computer have limitations in encouraging collaborative interactions and generalizing within-system interactions to real-world settings. Collaborative Virtual Environments (CVEs), which can support interactions between multiple real-users across the internet, offer a platform for collaborative interactions and natural communications between these users (M Schmidt et al. 2011). Therefore, we designed CVEs to understand and enhance collaborative interactions and communication of children with ASD in peer-mediated interactions.

A Collaborative Virtual Environment (CoMove)

In Chapter II, we present the design and application of a CVE system, named CoMove, to understand and encourage collaborative interactions and communication skills of children with ASD. In CoMove, we designed multiple collaborative puzzle games that were equipped with collaborative strategies in order to

encourage collaborations between participants. The collaborative strategies were implemented with a hybrid automata model to promote three kinds of collaborative interactions, i.e., turn-taking, information sharing, and simultaneous work (taking actions at the same time). These three kinds of collaborative interactions were selected since they related to a variety of real-world social communication skills (White et al. 2007). In addition, we provided an objective measurement method to measure the participants' communication and collaboration skills within CoMove.

A total of 28 children, 7 age-matched ASD/TD pairs and 7 age-matched TD/TD pairs (age range: 7 – 17 years) were recruited to participate in a feasibility study. Each pair of the participants stayed in two different rooms of the same building, and completed 17 collaborative puzzle games within CoMove. The feasibility study included pre- and post-tests to evaluate impacts of CoMove on the participants' collaborative interactions. A Wilcoxon Signed-rank test showed statistically significant improvements of the interactions in the post-test compared to the pre-test. These results indicated the feasibility of the CVE for the targeted population, and the potential of the objective measurement method in indexing important aspects of interactions in the CVE.

Collaborative Virtual Environment on the Android platform

Mobile applications have the potential to engage children with ASD (Tanaka et al. 2010) by creating ubiquitous learning environments (Gravenhorst et al. 2015). Therefore, we developed a CVE on the Android platform to understand their collaborative interactions and communication skills when they used the mobile application. One challenge of designing CVEs is to support face-to-face communication, including both verbal and non-verbal communications (Montoya et al. 2011; Laffey et al. 2009). We implemented both audio and video chat functionalities in the CVE on the Android platform to address this challenge. By following the same interaction protocol as that in the previous CVE study, five age- and gender-matched pairs of participants (age range: 7 – 17 years) were recruited in a preliminary study using the CVE on the Android platform. It was found that children with ASD had different performance regarding their verbal-communication patterns compared to their TD peers. In particular, each of the

children with ASD spoke fewer words, asked fewer questions, and gave more responses, than his/her TD partner in the pre-test. However, these differences were not statistically significant.

This work resulted in two published research manuscripts and one conditionally accepted manuscript:

Lian Zhang, Qiang Fu, Amy Swanson, Amy Weitlauf, Zachary Warren, and Nilanjan Sarkar, "Design and Evaluation of a Collaborative Virtual Environment (CoMove) for Autism Intervention", *ACM Transactions on Accessible Computing* (Accepted).

Lian Zhang, Zachary Warren, Amy Swanson, Amy Weitlauf, and Nilanjan Sarkar. "Understanding Performance and Verbal-Communication of Children with ASD in a Collaborative Virtual Environment." Journal of autism and developmental disorders (2018): 1-11.

Lian Zhang, Megan Gabriel-King, Zachary Armento, Miles Baer, Qiang Fu, Huan Zhao, Amy Swanson, Medha Sarkar, Zachary Warren, Nilanjan Sarkar. "Design of a Mobile Collaborative Virtual Environment for Autism Intervention." *International Conference on Universal Access in Human-Computer Interaction*. 2016.

1.4.2. An intelligent agent for measurements in a CVE

Although CVEs have the advantages to support realistic interactions and flexible communication between real-users, measuring the interactions is limited by two fundamental challenges. First, the dynamic social interactions within CVE systems are partner dependent. That is, interactions within the CVE change based on specific partner input. This fundamentally limits the ability to create consistent, controlled, and replicable interactions within the CVE. Second, open-ended CVE systems pose no restriction in verbal communication between users. As such, subsequent manual coding of interactions is necessary to understand patterns of communication for meaningful measurements and intervention. In order to address these challenges, we designed an intelligent agent that could play games and communicate with humans to automatically measure both communication and collaboration skills of the ASD population. Please note that designing a system that can understand unrestricted naturalistic conversation of humans (i.e., the Turing test) is still problematic from a technical point of view. However,

it is possible to understand controlled conversations in a narrowly defined domain using an intelligent agent. Our intelligent agent was designed to communicate and play collaborative games in a CVE.

In Chapter III, we present the design of the intelligent agent that could communicate and play collaborative games with children with ASD and their TD peers in order to automatically measure their behaviors in a CVE. This intelligent agent was developed using a novel hybrid method, which combined a dialogue act classification and a finite state machine. This method had the advantage to not only enable the agent's capability to communicate and play games, but also generate data for meaningful measurements of the interactions. A preliminary study involving five children with ASD (age range: 7 – 17 years) was conducted to test the feasibility of the intelligent agent. Results of this preliminary study indicated that the agent could i) properly initiate conversations with an accuracy of 82.93%, and ii) correctly respond to a human with an accuracy of 89.20%. These results were comparable with existing intelligent agents with conversation capabilities for TD population. The game performance of the participants, which was measured using a collaborative movement ratio, when they interacted with the agent in the preliminary study was comparable to the performance of participants in peer-mediated interactions in Chapter II. These results indicated the potential of the intelligent agent to communicate and play games in the CVE.

There is one research paper to be submitted on this study:

Lian Zhang, Amy Swanson, Amy Weitlauf, Zachary Warren, and Nilanjan Sarkar, "Design and Development of an Intelligent Agent to Measure Communication and Collaboration Skills in Collaborative Virtual Environment for Autism Intervention", IEEE transactions on learning technologies (to be submitted).

1.4.3. A framework to measure communication skills and collaboration skills

In order to address limitations in measuring peer-mediated interactions in CVEs, we applied the intelligent agent to control and index important aspects of the interactions in the CVE. The majority of existing CVEs for ASD intervention measure users' collaborative interactions and communication skills

based on a human-coding methodology (Matthew Schmidt et al. 2012). We developed a measurement system that applied the intelligent agent to automatically measure the users' both communication and collaboration skills. This system measured the skills in three steps. First, the system generated task-performance and verbal-communication features to represent the users' behaviors. Then, we evaluated whether the system could accurately generate these features, as well as whether these features could reflect important aspects of the user behaviors in the CVE using statistical tests. Finally, all of the features were combined together to measure users' both communication and collaboration skills with machine learning methods.

A total of 40 children, 20 age-, and sex-matched ASD/TD pairs (age range: 7 – 17 years), were recruited to participate in a feasibility study. Each pair of the participants played nine collaborative games. Two children of a pair, first, played a game with each other in a human-human interaction (HHI) mode, and then played the same game with their own intelligent agents in a human-agent interaction (HAI) mode. The HAIs were designed for the intelligent agent to control and measure the children's skills; while the HHIs were designed to provide ground truth to evaluate the measurements in the HAIs. We found strong correlations between some system-generated features and the communication skills of the participants in the HAIs, as well as strong correlations between some system-generated features and the collaboration skills in the HAIs. We also achieved high accuracies when measured both communication and the collaboration skills based on these system-generated features. These results indicated that the intelligent agent had the potential to automatically measure the participants' both communication and collaboration skills within the CVE. Additionally, we found strong correlations between some features in HAIs and the features in HHIs. These results indicated that intelligent agent-based interactions could reflect important aspects of the human-human interactions.

1.4.4. Multimodal fusion for cognitive load measurement

In this work, we designed an intelligent HCI system to measure cognitive load of individuals with ASD. Cognitive load is believed to be a crucial factor in how children with ASD acquire knowledge and

skills (Paas et al. 2003). However, individuals with ASD have difficulties in correctly recognizing their cognitive load (Rajendran and Mitchell 2007). Therefore, a way to automatically measure the cognitive load is needed. In this context, researchers have investigated methods to measure the cognitive load using eye gaze signals, peripheral physiology signals, and EEG signals, respectively. Combining these multimodal sensor signals may lead to a higher accuracy in the measurements.

We designed a framework to measure the cognitive load of individuals with ASD from the multi-model signals with data fusion technologies. We explored three data fusion strategies: feature-level fusion, decision-level fusion, and hybrid-level fusion. In addition, we developed a weight selection mechanism to compute parameters of the decision-level fusion. The weight selection mechanism reduced computational load but still generated the optimal weights comparing to the widely used exhaustive search method (Koelstra et al. 2012). Based on these data fusion strategies, we found that multimodal fusion outperformed single modality classification in measuring cognitive load of the individuals with ASD.

There are eight research papers published on this study:

Lian Zhang, Joshua Wade, Dayi Bian, Jing Fan, Amy Swanson, Amy Weitlauf, Zachary Warren, Nilanjan Sarkar, "Cognitive load measurement in a Virtual Reality-based Driving System for Autism Intervention", IEEE Transactions on Affective Computing, 2016.

Joshua Wade, **Lian Zhang**, Dayi Bian, Jing Fan, Amy Swanson, Amy Weitlauf, Medha Sarkar, Zachary Warren, Nilanjan Sarkar, "A Gaze-Contingent Adaptive Virtual Reality Driving Environment for Intervention in Individuals with Autism Spectrum Disorders", ACM Transactions on Interactive Intelligent Systems (2016).

Lian Zhang, Joshua Wade, Amy Swanson, Amy Weitlauf, Zachary Warren, and Nilanjan Sarkar. "Cognitive State Measurement from Eye Gaze Analysis in an Intelligent Virtual Reality Driving System for Autism Intervention", the sixth International Conference on Affective Computing and Intelligent Interaction (ACII2015)

Lian Zhang, Joshua Wade, Dayi Bian, Jing Fan, Amy Swanson, Amy Weitlauf, Zachary Warren, and Nilanjan Sarkar. “Multimodal Fusion for Cognitive Load Measurement in an Adaptive Virtual Reality Driving Task for Autism Intervention” The 17th International Conference on Human-Computer Interaction, 2015

Wade, Joshua, Dayi Bian, Jing Fan, **Lian Zhang**, Amy Swanson, Medha Sarkar, Amy Weitlauf, Zachary Warren, and Nilanjan Sarkar. "A virtual reality driving environment for training safe gaze patterns: application in individuals with ASD." In International Conference on Universal Access in Human-Computer Interaction, pp. 689-697. Springer International Publishing, 2015.

Lian Zhang, Joshua W. Wade, Dayi Bian, Amy Swanson, Zachary Warren, Nilanjan Sarkar, “Data Fusion for Difficulty Adjustment in an Adaptive Virtual Reality Game System for Autism Intervention“, in 16th International Conference on Human-Computer Interaction, 2014

Wade, Joshua, Dayi Bian, **Lian Zhang**, Amy Swanson, Medha Sarkar, Zachary Warren, and Nilanjan Sarkar. "Design of a virtual reality driving environment to assess performance of teenagers with ASD." In International Conference on Universal Access in Human-Computer Interaction, pp. 466-474. Springer International Publishing, 2014.

Bian, Dayi, Joshua W. Wade, **Lian Zhang**, Esubalew Bekele, Amy Swanson, Julie Ana Crittendon, Medha Sarkar, Zachary Warren, and Nilanjan Sarkar. "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, 2013.

The rest of this dissertation is organized as follows. In Chapter II, we present the design and application of a CVE, name CoMove, to encourage collaboration and communication between children with ASD and their TD peers. In Chapter III, we present the design of another CVE, i.e., a CVE on the Android platform, for the children to interact with each other using mobile devices. In Chapter IV, we present the design and development of an intelligent agent that could communicate and play games with the children in order to measure their behaviors within CoMove. Chapter V shows the application of the intelligent agent to measure both communication and collaboration skills of the children within the CVE.

Chapter VI discusses the development of another intelligent HCI system to measure cognitive load of the children using data fusion technologies. The final chapter summarizes the contributions of the current work and the possible future work.

1.5. References

Al-Omar, D., Al-Wabil, A., & Fawzi, M. Using pupil size variation during visual emotional stimulation in measuring affective states of non communicative individuals. In *International Conference on Universal Access in Human-Computer Interaction, 2013* (pp. 253-258): Springer

Anwar, A., Rahman, M. M., Ferdous, S., Anik, S. A., & Ahmed, S. I. A computer game based approach for increasing fluency in the speech of the autistic children. In *Advanced Learning Technologies (ICALT), 2011 11th IEEE international conference on, 2011* (pp. 17-18): IEEE

Association, A. P. (2000). *Diagnostic criteria from dsM-iV-tr*: American Psychiatric Pub.

Baecker, R. M. (2014). *Readings in Human-Computer Interaction: toward the year 2000*: Morgan Kaufmann.

Bauminger, N., Rogers, S., Aviezer, A., & Solomon, M. (2005). The friendship observation scale (FOS). *Unpublished manual, Bar Ilan University, Israel and University of California, Davis, CA*.

Bekele, E., Wade, J., Bian, D., Fan, J., Swanson, A., Warren, Z., et al. Multimodal adaptive social interaction in virtual environment (MASI-VR) for children with Autism spectrum disorders (ASD). In *2016 IEEE Virtual Reality (VR), 2016* (pp. 121-130): IEEE

Bekele, E., Wade, J. W., Bian, D., Zhang, L., Zheng, Z., Swanson, A., et al. Multimodal Interfaces and Sensory Fusion in VR for Social Interactions. In *International Conference on Virtual, Augmented and Mixed Reality, 2014* (pp. 14-24): Springer

Bekele, E., Young, M., Zheng, Z., Zhang, L., Swanson, A., Johnston, R., et al. A step towards adaptive multimodal virtual social interaction platform for children with autism. In *International Conference on Universal Access in Human-Computer Interaction, 2013* (pp. 464-473): Springer

Bellini, S., & Akullian, J. (2007). A meta-analysis of video modeling and video self-modeling interventions for children and adolescents with autism spectrum disorders. *Exceptional children, 73*(3),

264-287.

Ben-Sasson, A., Lamash, L., & Gal, E. (2013). To enforce or not to enforce? The use of collaborative interfaces to promote social skills in children with high functioning autism spectrum disorder. *Autism, 17*(5), 608-622.

Benford, S., Greenhalgh, C., Rodden, T., & Pycock, J. (2001). Collaborative virtual environments. *Communications of the ACM, 44*(7), 79-85.

Bernard-Opitz, V., Sriram, N., & Nakhoda-Sapuan, S. (2001). Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. *Journal of autism and developmental disorders, 31*(4), 377-384.

Bernard-Opitz, V., Sriram, N., & Sapuan, S. (1999). Enhancing vocal imitations in children with autism using the IBM speech viewer. *Autism, 3*(2), 131-147.

Bian, D., Wade, J., Swanson, A., Warren, Z., & Sarkar, N. Physiology-based affect recognition during driving in virtual environment for autism intervention. In *2nd international conference on physiological computing system (Accepted, 2015), 2015*

Bian, D., Wade, J., Warren, Z., & Sarkar, N. Online Engagement Detection and Task Adaptation in a Virtual Reality Based Driving Simulator for Autism Intervention. In *International Conference on Universal Access in Human-Computer Interaction, 2016* (pp. 538-547): Springer

Brusilovsky, P., & Millán, E. User models for adaptive hypermedia and adaptive educational systems. In *The adaptive web, 2007* (pp. 3-53): Springer-Verlag

Burke, M., Kraut, R., & Williams, D. Social use of computer-mediated communication by adults on the autism spectrum. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work, 2010* (pp. 425-434): ACM

Caltagirone, S., Keys, M., Schlief, B., & Willshire, M. J. (2002). Architecture for a massively multiplayer online role playing game engine. *Journal of Computing Sciences in Colleges, 18*(2), 105-116.

Cheng, Y., Chiang, H.-C., Ye, J., & Cheng, L.-h. (2010). Enhancing empathy instruction using a collaborative virtual learning environment for children with autistic spectrum conditions. *Computers & Education, 55*(4), 1449-1458.

Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early intensive behavioral treatment: Replication of the UCLA model in a community setting. *Journal of Developmental & Behavioral Pediatrics, 27*(2), S145-S155.

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.

Council, N. R. (1994). *Virtual reality: scientific and technological challenges*: National Academies Press.

Curtis, D. D., & Lawson, M. J. (2001). Exploring collaborative online learning. *Journal of Asynchronous learning networks, 5*(1), 21-34.

Developmental, D. M. N. S. Y., & Investigators, P. (2014). Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *Morbidity and mortality weekly report. Surveillance summaries (Washington, DC: 2002), 63*(2), 1.

Escobedo, L., Nguyen, D. H., Boyd, L., Hirano, S., Rangel, A., Garcia-Rosas, D., et al. MOSOCO: a mobile assistive tool to support children with autism practicing social skills in real-life situations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2012* (pp. 2589-2598): ACM

Gal, E., Bauminger, N., Goren-Bar, D., Pianesi, F., Stock, O., Zancanaro, M., et al. (2009). Enhancing social communication of children with high-functioning autism through a co-located interface. *Ai & Society, 24*(1), 75.

Golan, O., & Baron-Cohen, S. (2006). Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Development and psychopathology, 18*(02), 591-617.

Gravenhorst, F., Muaremi, A., Bardram, J., Grünerbl, A., Mayora, O., Wurzer, G., et al. (2015). Mobile phones as medical devices in mental disorder treatment: an overview. *Personal and Ubiquitous Computing, 19*(2), 335-353.

Heimann, M., Nelson, K. E., Tjus, T., & Gillberg, C. (1995). Increasing reading and communication skills in children with autism through an interactive multimedia computer program. *Journal of autism and developmental disorders, 25*(5), 459-480.

Hetzroni, O. E., & Tannous, J. (2004). Effects of a computer-based intervention program on the communicative functions of children with autism. *Journal of autism and developmental disorders, 34*(2), 95-113.

Hourcade, J. P., Bullock-Rest, N. E., & Hansen, T. E. (2012). Multitouch tablet applications and activities to enhance the social skills of children with autism spectrum disorders. *Personal and ubiquitous computing, 16*(2), 157-168.

Islam, M. M., Amin, M. A. B., & Biswas, P. (2013). Increasing Speech Ability of the Autistic Children by an Interactive Computer Game. *Global Journal of Computer Science and Technology, 13*(9).

Josman, N., Ben-Chaim, H. M., Friedrich, S., & Weiss, P. L. (2008). Effectiveness of virtual reality for teaching street-crossing skills to children and adolescents with autism. *International Journal on Disability and Human Development, 7*(1), 49-56.

Ketterl, M., Knipping, L., Ludwig, N., Mertens, R., Rahman, M., Ferdous, S., et al. (2011). Speech development of autistic children by interactive computer games. *Interactive Technology and Smart Education, 8*(4), 208-223.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2012). Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing, 3*(1), 18-31.

Konstantinidis, E. I., Luneski, A., Frantzidis, C. A., Costas, P., & Bamidis, P. D. A proposed framework of an interactive semi-virtual environment for enhanced education of children with autism spectrum disorders. In *computer-based medical systems, 2009. CBMS 2009. 22nd IEEE international symposium on, 2009* (pp. 1-6): IEEE

Kuriakose, S., & Lahiri, U. (2015). Understanding the Psycho-Physiological Implications of Interaction With a Virtual Reality-Based System in Adolescents With Autism: A Feasibility Study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 23*(4), 665-675.

Laffey, J., Schmidt, M., Stichter, J., Schmidt, C., & Goggins, S. iSocial: A 3D VLE for youth with autism. In *Proceedings of the 9th international conference on Computer supported collaborative learning-Volume 2, 2009* (pp. 112-114): International Society of the Learning Sciences

Lahiri, U., Bekele, E., Dohrmann, E., Warren, Z., & Sarkar, N. (2015). A physiologically informed virtual reality based social communication system for individuals with autism. *Journal of autism and developmental disorders, 45*(4), 919-931.

Lányi, C. S., & Tilinger, Á. Multimedia and virtual reality in the rehabilitation of autistic children. In *International Conference on Computers for Handicapped Persons, 2004* (pp. 22-28): Springer

Lee, Y.-C., Lee, J. D., & Boyle, L. N. (2007). Visual attention in driving: the effects of cognitive load and visual disruption. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 49*(4), 721-733.

Leman, P. J. (2015). How do groups work? Age differences in performance and the social outcomes of peer collaboration. *Cognitive science, 39*(4), 804-820.

Liu, C., Conn, K., Sarkar, N., & Stone, W. (2008). Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum Disorder. *International journal of human-computer studies, 66*(9), 662-677.

Livingstone, D., Kemp, J., & Edgar, E. (2008). From Multi-User Virtual Environment to 3D Virtual Learning Environment. *ALT-J: Research in Learning Technology, 16*(3), 139-150.

Millen, L., Cobb, S., Patel, H., & Glover, T. Collaborative virtual environment for conducting design sessions with students with autism spectrum conditions. In *Proc. 9th Intl Conf. on Disability, Virtual Reality and Assoc. Technologies, 2012* (pp. 269-278)

Mineo, B. A., Ziegler, W., Gill, S., & Salkin, D. (2009). Engagement with electronic screen media among students with autism spectrum disorders. *Journal of autism and developmental disorders, 39*(1), 172-187.

Montoya, M. M., Massey, A. P., & Lockwood, N. S. (2011). 3D collaborative virtual environments: exploring the link between collaborative behaviors and team performance. *Decision Sciences, 42*(2), 451-476.

Moore, D., Cheng, Y., McGrath, P., & Powell, N. J. (2005). Collaborative virtual environment technology for people with autism. *Focus on Autism and Other Developmental Disabilities, 20*(4), 231-243.

Moore, D., & Taylor, J. (2000). Interactive multimedia systems for students with autism. *Journal of Educational Media, 25*(3), 169-177.

Neumann, D., Spezio, M. L., Piven, J., & Adolphs, R. (2006). Looking you in the mouth: abnormal gaze in autism resulting from impaired top-down modulation of visual attention. *Social cognitive and affective neuroscience, 1*(3), 194-202.

Novak, D., Mihelj, M., & Munih, M. (2012). A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing. *Interacting with computers, 24*(3), 154-172.

Nuske, H. J., Vivanti, G., & Dissanayake, C. (2014). Brief report: evidence for normative resting-state physiology in autism. *Journal of autism and developmental disorders, 44*(8), 2057-2063.

Ozonoff, S., & Strayer, D. L. (2001). Further evidence of intact working memory in autism. *Journal of autism and developmental disorders, 31*(3), 257-263.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist, 38*(1), 63-71.

Parsons, S., & Cobb, S. (2011). State-of-the-art of virtual reality technologies for children on the autism spectrum. *European Journal of Special Needs Education, 26*(3), 355-366.

Parsons, S., Mitchell, P., & Leonard, A. (2004). The use and understanding of virtual environments by adolescents with autistic spectrum disorders. *Journal of autism and developmental disorders, 34*(4), 449-466.

Peacock, G., Amendah, D., Ouyang, L., & Grosse, S. D. (2012). Autism spectrum disorders and health care expenditures: the effects of co-occurring conditions. *Journal of Developmental & Behavioral Pediatrics, 33*(1), 2-8.

Pennington, R. C. (2010). Computer-assisted instruction for teaching academic skills to students with autism spectrum disorders: A review of literature. *Focus on Autism and Other Developmental Disabilities, 25*(4), 239-248.

Ploog, B. O., Banerjee, S., & Brooks, P. J. (2009). Attention to prosody (intonation) and content in children with autism and in typical children using spoken sentences in a computer game. *Research in Autism Spectrum Disorders, 3*(3), 743-758.

Rajendran, G., & Mitchell, P. (2007). Cognitive theories of autism. *Developmental Review, 27*(2), 224-

260.

Ramdoss, S., Machalicek, W., Rispoli, M., Mulloy, A., Lang, R., & O'Reilly, M. (2012). Computer-based interventions to improve social and emotional skills in individuals with autism spectrum disorders: A systematic review. *Developmental neurorehabilitation, 15*(2), 119-135.

Reimer, B., Fried, R., Mehler, B., Joshi, G., Bolfek, A., Godfrey, K. M., et al. (2013). Brief report: Examining driving behavior in young adults with high functioning autism spectrum disorders: A pilot study using a driving simulation paradigm. *Journal of autism and developmental disorders, 43*(9), 2211-2217.

Reynolds, S., Bendixen, R. M., Lawrence, T., & Lane, S. J. (2011). A pilot study examining activity participation, sensory responsiveness, and competence in children with high functioning autism spectrum disorder. *Journal of autism and developmental disorders, 41*(11), 1496-1506.

Rheingold, H. (1991). *Virtual Reality: Exploring the Brave New Technologies*: Simon & Schuster Adult Publishing Group.

Rogers, S. J. (1998). Empirically supported comprehensive treatments for young children with autism. *Journal of clinical child psychology, 27*(2), 168-179.

Rutherford, M., & Towns, A. M. (2008). Scan path differences and similarities during emotion perception in those with and without autism spectrum disorders. *Journal of autism and developmental disorders, 38*(7), 1371-1381.

Schmidt, M., Laffey, J., & Stichter, J. Virtual social competence instruction for individuals with autism spectrum disorders: Beyond the single-user experience. In *Proceedings of CSCL, 2011* (pp. 816-820)

Schmidt, M., Laffey, J. M., Schmidt, C. T., Wang, X., & Stichter, J. (2012). Developing methods for understanding social behavior in a 3D virtual learning environment. *Computers in Human Behavior, 28*(2), 405-413.

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.

Stone, W. L., Ousley, O. Y., Yoder, P. J., Hogan, K. L., & Hepburn, S. L. (1997). Nonverbal

communication in two-and three-year-old children with autism. *Journal of autism and developmental disorders, 27*(6), 677-696.

Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*: Cambridge university press.

Tanaka, J. W., Wolf, J. M., Klaiman, C., Koenig, K., Cockburn, J., Herlihy, L., et al. (2010). Using computerized games to teach face recognition skills to children with autism spectrum disorder: the Let's Face It! program. *Journal of Child Psychology and Psychiatry, 51*(8), 944-952.

Wade, J., Zhang, L., Bian, D., Fan, J., Swanson, A., Weitlauf, A., et al. (2016). A Gaze-Contingent Adaptive Virtual Reality Driving Environment for Intervention in Individuals with Autism Spectrum Disorders. *ACM Transactions on Interactive Intelligent Systems (TiiS), 6*(1), 3.

Wainer, A. L., & Ingersoll, B. R. (2011). The use of innovative computer technology for teaching social communication to individuals with autism spectrum disorders. *Research in Autism Spectrum Disorders, 5*(1), 96-107.

Wang, M., & Reid, D. (2010). Virtual reality in pediatric neurorehabilitation: attention deficit hyperactivity disorder, autism and cerebral palsy. *Neuroepidemiology, 36*(1), 2-18.

White, S. W., Keonig, K., & Scahill, L. (2007). Social skills development in children with autism spectrum disorders: A review of the intervention research. *Journal of autism and developmental disorders, 37*(10), 1858-1868.

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.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology, 4*(3-4), 129-164.

Xu, D., & Wang, H. (2006). Intelligent agent supported personalization for virtual learning environments. *Decision Support Systems, 42*(2), 825-843.

Yamamoto, J.-i., & Miya, T. (1999). Acquisition and transfer of sentence construction in autistic students:

Analysis by computer-based teaching. *Research in Developmental Disabilities, 20*(5), 355-377.

Zancanaro, M., Pianesi, F., Stock, O., Venuti, P., Cappelletti, A., Iandolo, G., et al. (2007). Children in the museum: an environment for collaborative storytelling. In *PEACH-Intelligent Interfaces for Museum Visits* (pp. 165-184): Springer.

CHAPTER II. A COLLABORATIVE VIRTUAL ENVIRONMENT (COMOVE)

2.1. Abstract

A CVE, which is a computer-based, distributed, virtual space for multiple users to interact with one another and/or with virtual items, has the potential to support flexible, safe and peer-based social interactions. In this chapter, we presented the design of a CVE system, called CoMove, with the ultimate goals of measuring and potentially enhancing collaborative interactions and verbal-communication of children with ASD when they play collaborative puzzle games with their typically developing (TD) peers in remote locations. CoMove has two distinguishing characteristics: i) the ability to promote important collaborative behaviors (including information-sharing, sequential interactions, and simultaneous interactions) and to provide real-time feedback based on users' game performance; as well as ii) an objective way to measure and index important aspects of collaboration and verbal-communication skills during system interaction. A feasibility study with 14 pairs, 7 ASD/TD pairs and 7 TD/TD pairs, was conducted to initially test the feasibility of CoMove. Results of the study validated the system feasibility and suggested its potential to index important aspects of collaboration and verbal-communication.

2.2. Introduction

A collaborative virtual environment (CVE) is a computer-based, distributed, virtual space for multiple individuals to interact with one another and/or with virtual items (Benford et al. 2001). Compared to existing studies of interaction between children with ASD and computer-controlled virtual avatars (Moore et al. 2005; Cheng et al. 2010), multi-user CVEs present the opportunity for dynamic user-to-user interactions, instead of user-machine interactions, in shared virtual environments. Such systems, which are suited for collaboration and group work among real users, may offer an effective and beneficial way to foster social relationships among children with ASD and their typically developing (TD) peers (Leman 2015; Reynolds et al. 2011).

Some researchers have evaluated the user-to-user interactions using co-located devices, such as tabletop (Battocchi et al. 2009), tablet (Ben-Sasson et al. 2013; Hourcade et al. 2012), iPad (Boyd et al. 2015), and wearable devices (Boyd et al. 2016). These systems allow multiple users to share the same device for face-to-face interactions. However, face-to-face interactions may be initially difficult for children with ASD given the potential multisensory integration deficits associated with ASD (Ringland et al. 2016). Impairments in multisensory integration, which generally involves how information from the different sensory modalities, such as sight, sound, touch, smell, self-motion and taste, may be integrated by the nervous system (Stein et al. 2009), is one of the diagnostic criteria for ASD (McPartland et al. 2012). In addition, it is difficult to identify users using the co-located devices from a technical point of view(Boyd et al. 2015). As a result, one user may do all the tasks without collaboration with other users. A distributed system, which allows interactions among users from different locations and reduces information of some sensory modalities, may address these limitations (M Schmidt et al. 2011; Millen et al. 2012). In what follows, we reviewed previous literature in distributed systems to support user-to-user interactions for ASD intervention.

2.2.1. Related work

Massively multiplayer online games and social networks have been investigated for ASD intervention (Caltagirone et al. 2002; Livingstone et al. 2008). For example, Burke and colleagues analyzed the needs and effects of social communication (such as text-messaging, email, and Facebook) for adults with ASD (Burke et al. 2010). They reported many benefits, such as reduced stress and increased greetings. Ringland and colleagues applied the Minecraft game in order to see how individuals with ASD engaged in social play (Ringland et al. 2016). They found that individuals with ASD were as social as TD individuals in the game. Other multiplayer online games, such as Second-life(Newbutt)and Zody's world (Boyd et al. 2015), have also been successfully used to study behaviors of children with ASD in group work. However, researchers usually had limited access to the source code of these commercial games. As a result, it could be difficult to alter the games in a way that would allow researchers to structure and

investigate specific collaborative activities. In addition, most of these studies analyzed users' behaviors in games using observation-based methods. It will be important to explore other methods that can be used to quantitatively measure interactions between users in these games.

In this context, CVEs with specific social activities have designed and developed for ASD interventions. Cheng and colleagues designed a virtual restaurant in order to understand empathy in children with ASD (Cheng et al. 2010). They also developed a CVE with two other social scenes, a classroom scene and an outdoor scene, in order to promote social competence in the ASD population (Cheng and Ye 2010). 3D virtual avatars, which had gestures and facial expressions, were used as representations of real users in the environment, and were applied to investigate the ability of children with ASD to understand these social cues (i.e. gestures and facial expressions) in the environment. ISocial is another important CVE that was designed to investigate social competency in children with ASD (Laffey et al. 2012). In iSocial, individuals with ASD could interact with each other through the internet using their own 3D virtual avatars. Naturalistic practice learning activities (Wang et al. 2016) and a social competence curriculum (e.g., facial expression recognition) (Stichter et al. 2014) have been designed as activities in iSocial for understanding social competence. Finally, Wallace and colleagues designed a CVE to teach children with ASD greeting behaviors in a virtual gallery (Wallace et al. 2015). They found that children with ASD were less sensitive to a negative greeting from the human avatar in the virtual gallery than their TD peers. Although virtual avatars have been successfully used to represent users in these virtual environments (E. S. Liu and Theodoropoulos 2014), the efficacy of the use of avatars for presentation is under debate (Benford et al. 2001; Natkin and Yan 2006). Please also note that all the activities in these CVEs rely upon specific social skills and are designed to promote external goals of their respective curricula, rather than provide users with game tasks that require and reinforce efficient social interaction as part of the game environment itself.

In addition to these specified social curriculum activities, researchers have also designed collaborative games as interactive activities in CVEs for ASD intervention. Millen and colleagues have developed a CVE with a *block party* game, which requires users to select the same blocks in order to build a tower,

and a *talk about CVE* game, which was designed for children with ASD to practice conversation skills (Millen et al. 2011). However, these games did not require users to take actions at the same time (i.e., simultaneous interactions). This study reported preliminary results of a self-report questionnaire, which showed that children with ASD had improved engagement in these CVE-based collaborative games (P. L. Weiss et al. 2011). Schmidt and colleagues have proposed a game-based learning environment for individuals with ASD to learn computational thinking and social skills in groups across the internet (Matthew Schmidt and Beck 2016). In this environment, users would be able to collaborate with each other and play Minecraft[1] videogame according to specific collaboration rules. At the time of writing this article, however, results had not been reported.

Although existing massively multiplayer online games and CVEs, which could support flexible interactions between real users from different locations, have been successfully applied for individuals with ASD to practice specific social skills, most of them were not designed to facilitate collaboration, such as taking actions at the same time (Benford et al. 2001). Designing systems to facilitate collaboration has two primary challenges. The first challenge is related to the design of collaborative activities themselves in order to foster collaboration. In other words, structuring interactions within a game to require users take collaborative actions, such as taking actions at the same time, sharing information with each other, and playing in order. Collaboration is not something that simply happens whenever users come together (Dillenbourg 2002). Therefore, carefully-designed collaboration strategies are needed in order to enable and encourage collaboration in the environment. The second challenge is related to evaluating interactions to understand users' collaborative behaviors in specific CVEs and/or validate effects of the CVEs on the users' collaboration and communication skills. Evaluating interactions is challenging, given the unrestricted conversations and complex interactions between real users.

A potential way to address the first challenge in this area is to design collaborative games with embedded collaboration strategies to promote collaboration. A few studies have sought to promote

[1] https://education.minecraft.net/

collaboration in the ASD population by designing collaborative games and equipping these games with collaborative strategies using co-located multi-touch devices (Nakano et al. 2011; Noor et al. 2012). Gel and colleagues have designed and developed shared storytelling games on co-located multi-touch devices to promote collaboration (Bauminger et al. 2007). These games could not only support individual actions but also require simultaneous actions through a collaboration strategy, which require two users to touch and drag items simultaneously in order to move them (Cappelletti et al. 2004). Gel and colleagues found that participants initiated more positive social interactions, had more shared play, and performed fewer autistic behaviors while playing these games (Gal et al. 2009). Their study also indicated that children with ASD had more positive social interactions and collaborative play in the games with the collaboration strategy than in free-play conditions (Ben-Sasson et al. 2013). Battocchi and colleagues designed collaborative puzzle games with a similar collaboration strategy (Battocchi et al. 2009). They also demonstrated that this kind of puzzle game had positive effects on collaboration in children with ASD. These studies have shown that collaborative games equipped with deliberate collaboration strategies have the potential to promote collaboration in children with ASD. However, these games were developed using co-located multi-touch devices instead of distributed CVEs. Co-located multi-user systems can support face-to-face communication with both verbal and non-verbal cues, while communications in distributed CVEs often lack nonverbal cues (Montoya et al. 2011). Therefore, these existing collaboration strategies cannot be directly used in distributed CVE systems.

In order to address the first challenge, we have developed novel collaborative games with collaborative strategies to promote three important collaborative behaviors, i.e., information-sharing, sequential work, and simultaneous work (Rummel and Spada 2005; Johnson and Johnson 1996; Gal et al. 2005b). These collaborative behaviors (i.e., sequential work, information sharing, and simultaneous work) were targeted in this study because they relate to a variety of social settings, including employment, education, and game play(White et al. 2007). Specifically, in group work, people need to be able to take turns (sequential work), decide on certain aspect of the work and then effectively deliver relevant information (sharing information), and conduct tasks collaboratively (simultaneous work)(Leman 2015). Sequential work

implies that users take actions one by one (turn taking). Turn taking is a life skill necessary for social success in all environments (Rao et al. 2008; Bernard-Opitz et al. 2001). Information-sharing enables individuals to share resources and knowledge in order to achieve the same goal (Johnson and Johnson 1996). Weiss and colleagues observed that information sharing is one of the best ways to train social skills in children with ASD (M. J. Weiss and Harris 2001). Information sharing is also important for children with ASD to build friendships with others (Rao et al. 2008). Another aspect of group work, i.e., simultaneous work, requires that members of a group work together at the same time (Leman 2015; Gal et al. 2005a). It has been found that the simultaneous interactions could improve social skills by fostering the recognition of the presence of the other, and enhancing interest in partners (Gal et al. 2009; Zancanaro et al. 2007).

The second challenge in this area is to efficiently measure interactions to understand users' within system behaviors as well as evaluate generalized improvements beyond skill systems and training programs (Anagnostou et al. 2015). In order to understand interactions between multiple users, several methods, such as self-report, interviews, observations, performance, and dialogue analysis, have been explored by researchers in multi-user systems (Gress et al. 2010). Cheng and colleagues used a self-report method and an observation method to evaluate social competence of children with ASD in their CVE (Cheng and Ye 2010). In iSocial (Matthew Schmidt et al. 2012; Stichter et al. 2014), the authors analyzed users' behavior by coding their reciprocal social interactions (e.g., conversation initiations, responses and continuations), identifying their use of available avatar-based gestures and movement, and rating their behaviors. However, some aspects of these measurements are task-dependent. Therefore, the methods used to measure social activities in these CVEs cannot be directly applied to evaluate collaborative interactions in our CVE systems. In this Chapter, we have adapted these existing methods to develop a set of metrics for objective measurement of interactions in our CVE-based collaborative games.

2.2.2. Current work

The primary contributions of this chapter are: i) designing a novel computer-based CVE, named CoMove, that can promote important collaborative behaviors, such as sequential work, information sharing, and simultaneous work, between two users in a flexible manner, as well as provide appropriate real-time feedback based on their performance; ii) providing a potential way to objectively measure both collaborative and communicative behaviors of the users when they play these collaborative games; and iii) presenting the results of a feasibility study involving 7 pairs of ASD/TD and 7 pairs of TD/TD children to test system feasibility, and to assess the capacity of CoMove to index important aspects of interactions in the system.

CoMove was designed with the ultimate goals of measuring and potentially enhancing collaborative interactions and verbal-communication of children with ASD. In this study, we tested the feasibility of CoMove for children to collaboratively interact and communicate with each other, as well as its potential to index important aspects of the interactions. Testing the system feasibility and its measurement capability lays the groundwork for future investigations into how changes within the system may generalize to real world interactions. We will evaluate effects of CoMove on collaborative and verbal-communication skills of children with ASD in their real world interactions in the future.

The rest of the chapter is organized as follows. Section 2.3 presents the development of CoMove with an emphasis on designing collaborative puzzle games to promote collaboration, and generating collaboration- and communication-related data to objectively measure their behaviors. Section 2.4 provides information about the tasks, participants and experimental protocol. The results and discussion are presented in Section 2.5. Finally, Section 2.6 summarizes the contributions of this chapter, discusses limitations of the current work, and indicates potential future improvements.

2.3. System Design

CoMove is a distributed virtual environment system for two users to communicate and play collaborative puzzle games. The collaborative games in CoMove were designed to promote collaboration

through collaborative strategies and to provide feedback based on users' performance. In addition, an objective measurement method was developed for understanding users' collaboration and communication skills within the system. In what follows, we describe i) the architecture of CoMove, ii) the characteristics of these collaborative games, and iii) the details of the objective measurement approach.

2.3.1. Architecture of CoMove

CoMove was designed for two geographically distributed users using their own computers, where each computer is a node of the CVE system, to communicate and collaborate in a shared environment. Given this goal, we designed the architecture shown in Fig. 1, which is divided into a system architecture and an application architecture. The system architecture shows how two nodes (users' computers) are connected and how application data are distributed, while the application architecture is composed of the components used to implement functionalities of the application.

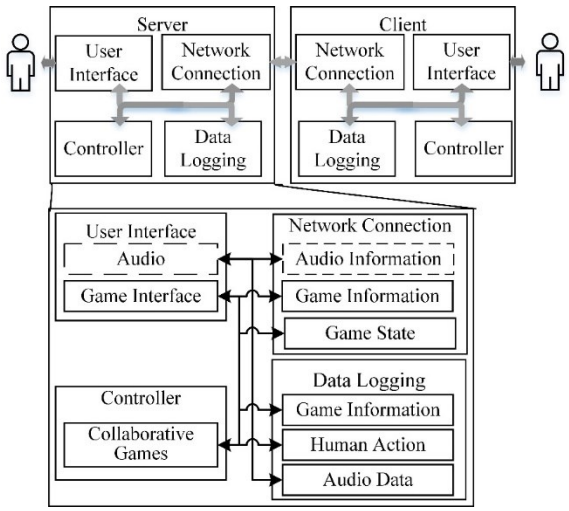

Fig. 1 The architecture of CoMove

Generally, the design of a CVE system involves i) how nodes (users' computers) are connected, and ii) how application data are distributed (Fleury et al. 2010). The design of the system architecture of CoMove aimed to provide an efficient way to address these two issues given specific system requirements. We selected a server-client model for node connection because it is simpler to maintain state consistency compared to the peer-to-peer model (Macedonia and Zyda 1997). The application data

were distributed with a replicated architecture, which replicated the entire application data on each node of CoMove. This replicated architecture can improve the use of network resources since the display data do not need to be transmitted over the network using this architecture (Suthers 2001). Usually, a scalable CVE system uses a separate central server to store all the application data, distribute specific data to each node, and update games states of all the nodes (Gautier and Diot 1998). Since CoMove targets interactions between two players, no separate central server is necessary at this point. In CoMove, one node is the server, another node is the client, and both nodes have the entire application data. This kind of architecture is convenient and has low network load for a two-player CVE system.

Generally, an application architecture is designed to enable specific functionalities. The application architecture of CoMove was designed to i) enable communication and game playing between two geographically distributed users, and ii) understand their behaviors in the system. The application architecture of CoMove has four components: an interface component, a controller component, a network connection component, and a data logging component. An interface component is generally used to get users' inputs and execute an application's outputs, while a controller component is often used to make decisions and generate responses in computer-based systems (Dix 2009). In order to support both verbal-communication and game-playing, the interface component of CoMove could support both audio input/output (through a microphone and a speaker) and game-related input/output (using a mouse and a graphic display monitor). The controller component of CoMove was implemented using a Finite State Machine (FSM) with the objective to i) manage multiple collaborative games to facilitate sequential interactions, information sharing, and simultaneous interactions, and ii) provide appropriate performance-based feedback to enhance learning. A network connection component is usually used by a CVE system to connect its multiple nodes (Bowers et al. 1996). We selected a server-client model to connect distributed nodes for simplicity, as discussed earlier. A data logging component (called "data collection" in iSocial (Matthew Schmidt et al. 2012)) is commonly used in CVEs for ASD intervention in order to understand behaviors from the logged data since understanding behaviors is one of the primarily goals in this kind of system. The data logging component of CoMove could log game information, human actions,

and audio data to understand children with ASD's collaborative performance and communication skills in collaborative games. In what follows, we present in detail the development of the controller, network connection, and data logging components of CoMove.

2.3.1. Controller

The controller component of CoMove was developed in order to manage multiple collaborative puzzle games. We selected collaborative puzzle games as the collaborative activities in CoMove because these games have been widely accepted as engaging children with ASD in collaborative interactions (Battocchi et al. 2010), and they are suitable for implementing multiple collaborative strategies to promote collaboration (Cappelletti et al. 2004). The logic to manage these collaborative puzzle games is modeled using a FSM with hierarchy and concurrency, as shown in Fig. 2.

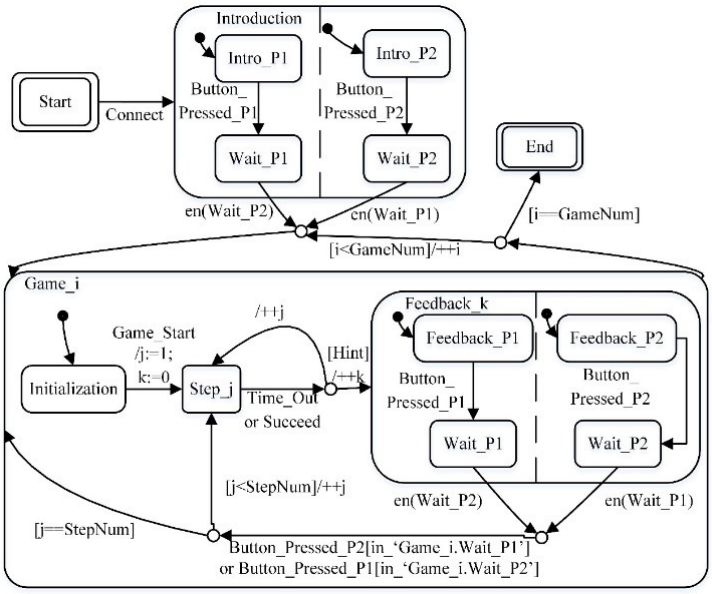

Fig. 2 Finite state machine in the controller component

The *Start* state of the FSM is the start point of the system. Two players will start their applications in this state and log into the shared environment from their own nodes. Subsequently, an introduction on how to use the environment and how to play the collaborative puzzle games is provided to these players in the *Introduction* state of the FSM. The *Introduction* state is a concurrent state, which can not only enable two players to read game information with different reading speeds but also synchronize their

states to start a game at the same time. We used a hierarchical state, the *Game_i* state, to present a collaborative puzzle game. The *Game_i* state has multiple sub-states, such as the *Initialization* state, the *Step_j* state and the *Feedback_k* state. The system starts a new game when an old one ends. This logic enables players to play multiple collaborative puzzle games that require multiple kinds of collaborative behaviors, i.e., sequential work, information sharing, and simultaneous work. The system can also offer performance-based feedback in the *Feedback_k* state to help players complete the games.

2.3.1.2. Network connection

The network connect component is mainly used to transfer data between different nodes. In CoMove, we designed an efficient data transmission mechanism that used different strategies to transfer different types of data. The trivial game information, such as puzzle piece position, is transferred between the two nodes without synchronization. The game state information is transferred using a hand-shaking mechanism for a stable synchronization. The server node of CoMove executes all the computations of the system, such as computing the game states. These computed game states are transferred in 7 steps: 1) the client node sends its new data to the server node (this step can be skipped if the new data occurs at the server node); 2) the server node then computes the new game state; 3) and sends the new game state as a synchronization request to the client node; 4) the client node updates its game state; 5) and sends a synchronization acknowledgement to the server node; 6) after receiving the synchronization acknowledgement from the client node, the server node updates its state; and 7) sends back an acknowledgement to the client node. This hand-shaking mechanism guarantees synchronization between two nodes. The audio data transmission is implemented using the Skype [2] software for simplicity and stability.

2.3.1.3. Data Logging

The data logging component of CoMove is used to store the performance- and communication-related data in order to understand the corresponding within-environment behaviors. The recorded data include

[2] skype.com

game information data as well as player behaviors and audio data during game play. These data are recorded in real-time with time stamps, which can be used for offline synchronization.

With these components, CoMove has the capacity to enable collaboration and communication between two geographically distributed players and record data to understand their behaviors in the system. The collaborative puzzle games in CoMove were designed with embedded strategies to promote collaboration. A framework for objective measurement of interactions in CoMove was designed in order to understand important aspects of the within-system behaviors. In what follows, we describe the collaborative puzzle games and the objective measurement framework.

2.3.2. Collaborative puzzle games

The puzzle games were designed to promote collaboration in the CVE for users with ASD. Three important collaborative behaviors, i.e., sequential work, information-sharing, and simultaneous work, were evaluated. In order to evaluate these three collaborative behaviors, three types of games, turn-taking games, information-sharing games, and collaboration games, were designed.

In each game, players were required to assemble a specific shape by dragging several puzzle pieces following specific rules. In the turn-taking games, each player had full control over the puzzle pieces during his/her turns. In the information-sharing games, colors of some puzzle pieces were hidden for one player while they were visible to the other player. Therefore, the players needed to ask and share the color information in order to move the correct pieces in this type of game. Finally, the collaboration games were implemented with a joint play strategy, which requires two players to drag a puzzle piece in the same direction simultaneously in order to move it. These three types of games together require sequential interactions, simultaneous interactions, and sharing of information. In addition, the system can provide performance-based feedback in these games to help users complete the games. The collaborative puzzle games developed in CoMove were composed of multiple tangram games (Fig. **3** (P1_1) and (P2_1) shows an instance of tangram games) and a castle-building game (shown in Fig. **3** (P1_3) and (P2_3)).

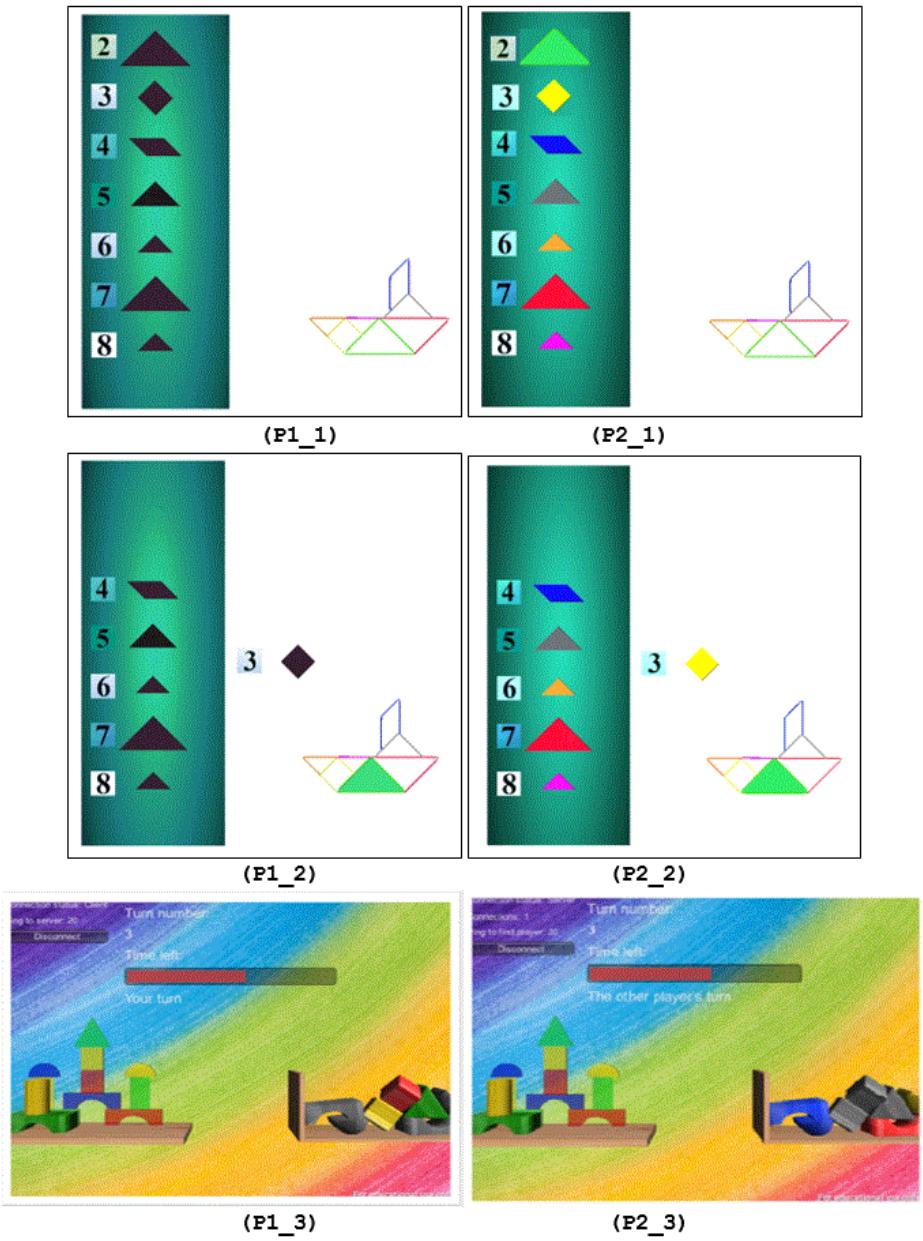

Fig. 3 P1_1 and P2_1 are Screenshots of a tangram game captured from two CVE nodes; P1_2 and P2_2 are Screenshots of the same tangram game from the nodes after two pieces being moved; and P1_3 and P2_3 are screenshots of a castle game of the nodes

Table 1 The value of the configuration features for each game

Game name	Who can see the colors of all the puzzle pieces	Who can move the puzzle pieces	Who can rotate all the puzzle pieces	Time duration (seconds)
T1	P1* and P2*	P1 in step 1, 3, 5, 7; P2 in step 2, 4,6	Auto**	30
T2	P1 and P2	P1 in step 3, 4, 7; P2 in step 1, 2, 5, 6	Auto	30
T3	P1 and P2	P1 and P2 in all steps	Auto	45
T4	P1	P2 in all steps	P2	40
T5	P2	P1 in all steps	P1	40
T6	P1	P1 and P2 in all steps	P2	50
T7	P2	P1 and P2 in all steps	P1	50

**P1* means the first player, and *P2* means the second player

** *Auto* means the puzzle pieces will be rotated automatically by the system

In each tangram game, players were required to assemble a specific shape from seven flat pieces. Each game had seven steps. In each step, a single puzzle piece needed to be moved to its target position. A total of seven tangram games were designed with five configuration features: 1) who can see the colors of all the puzzle piece; 2) who can move puzzle pieces; 3) who can rotate puzzle pieces; 4) the maximum time duration of a step; and 5) the feedback information. The values of the first four features of each game are listed in Table 1. For example, in game T1, both P1 and P2 can see the color of puzzle pieces. In this game, P1 and P2 could move puzzle pieces one by one, and the puzzle pieces could rotate automatically. The maximum time duration for a player to successfully move a puzzle piece in game T1 is 30 seconds. If the player failed to move any puzzle piece within the time duration, the system automatically moved a puzzle piece to its target position. The fifth feature, i.e., feedback information, is about how to move or rotate pieces in the game. For example, an example of feedback information in the collaboration game, T3, is "*Maybe two people are needed to move this puzzle piece".* In a tangram game, feedback information is offered when players fail the first two steps and the first four steps of the game.

The main focus of these games is to facilitate information sharing as well as promote both sequential and simultaneous interactions. The information-sharing was facilitated by hiding information for one of the two users. For example, in game T4, the first player can see colors of all puzzle pieces, while the

second player can move all the pieces. Therefore, the first player needs to share color-information with the second player in order for the second player to move the correct puzzle pieces.

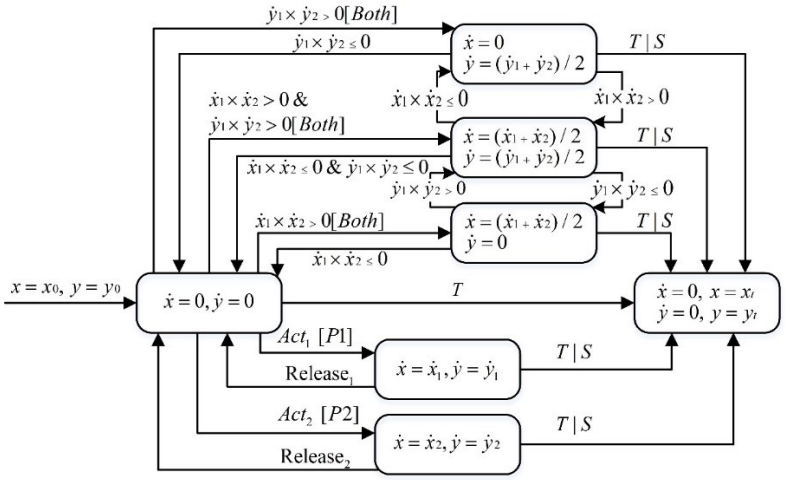

Fig. 4 The hybrid automaton

In order to promote both sequential and simultaneous interaction, we developed a collaboration strategy, which requires two players to move puzzle pieces individually or simultaneously in order to complete a specific game. Previous studies using multi-touch devices for co-located multi-user interactions implemented the collaboration strategy by defining rules, such as holding on the same puzzle piece (Gal et al. 2009; Fan et al. 2014). Those co-located systems enable players to communicate with each other in a face-to-face mode, which makes it easy to follow these rules. It is challenging to implement natural face-to-face communication in distributed CVE systems from a technical point of view (Montoya et al. 2011). Interactions in distributed systems lack natural non-verbal cues (such as gestures and eye contacts). This fact makes it harder for users to follow these rules in distributed CVE systems. Therefore, we developed a different way to implement this collaboration strategy in our distributed CVE system in order to fulfill the requirements to promote both sequential and simultaneous interactions.

Table 2 Symbols used in the hybrid automaton and their descriptions

Symbol	Type	Description
Act_i, i=1,2	Event	The first (i=1) or the second (i=2) player takes actions on the puzzle piece
$Release_i$, i=1,2	Event	The first (i=1) or the second (i=2) player release the puzzle piece
P_i, i=1,2	Condition	The first (i=1) or the second (i=2) player's turn to move puzzle pieces (turn-taking games)
Both	Condition	Two players move puzzle pieces together (collaboration games)
T	Event	Time out
S	Event	Succeed
x	Variable	Shift of a puzzle piece in the horizontal direction
y	Variable	Shift of a puzzle piece in the vertical direction
x_0	Variable	Initial position of a puzzle piece in the horizontal direction
y_0	Variable	Initial position of a puzzle piece in the vertical direction
x_t	Variable	Target position of a puzzle piece in the horizontal direction
y_t	Variable	Target position of a puzzle piece in the vertical direction
$\dot{x}$	Variable	Moving speed of a puzzle piece in the horizontal direction
$\dot{y}$	Variable	Moving speed of a puzzle piece in the vertical direction
$\dot{x}_i$, i=1, 2	Variable	Dragging speed of a puzzle piece by the ith player in the horizontal direction
$\dot{y}_i$, i=1, 2	Variable	Dragging speed of a puzzle piece by the ith player in the vertical direction

The hybrid automaton is used to represent how a puzzle piece responds when two players take actions in different games. In a collaboration game, the condition *Both* is true in the hybrid automaton. Under this condition, for example, if two players drag the puzzle piece towards the same direction ($\dot{x_1} \times \dot{x_2} > 0$), the velocity of the puzzle piece in the horizontal direction is $(\dot{x_1} + \dot{x_2})/2$. If one player drags the piece to right and the other player drag it to left ($\dot{x_1} \times \dot{x_2} < 0$) or one player stops dragging in the horizontal direction($\dot{x_1} \times \dot{x_2} = 0$), the velocity of the puzzle piece in the horizontal direction is 0. When the puzzle piece is successfully moved to its target area (presented with the event S) or the time reaches the maximum time duration (presented with the event T), the position of the puzzle piece is automatically set to (x_t, y_t). Using this hybrid automaton model, players are forced to interact differently, i.e., dragging individually or dragging simultaneously, in different games.

The castle-building game, as shown in Fig. 3 P1_3 and P2_3, is another CoMove game, which can enable turn-taking and information sharing as well provide performance-based feedback. The game is also designed based on the five configuration features mentioned in the tangram game. Each player could only see the colors of five pieces and could only move the other five pieces, which are summarized in Table 3 using the *i) who can move the puzzle pieces* and *ii) who can see the colors of the puzzle pieces*. P1 in Table 3 means the first player and P2 means the second player. The index of each puzzle piece is shown in Fig. 5. These puzzle pieces can be automatically rotated when they are correctly located. The maximum time duration without successfully moving a puzzle piece is 30 seconds. The castle game was different from the tangram games since its pieces had gravity and therefore needed to be built in order. In summary, players can move the pieces in turns with order constraints in the castle game. A player may have no piece to move during his/her turn because of the order constraints. Thus, a feedback is displayed for the player to skip his/her turn when no piece is movable for the player.

Table 3 The game information of the castle game

Index of a puzzle piece	1	2	3	4	5	6	7	8	9	10
Who can move the puzzle pieces	P1	P1	P1	P2	P2	P1	P1	P2	P2	P2
Who can see the color of the puzzle pieces	P2	P2	P2	P1	P1	P2	P2	P1	P1	P1

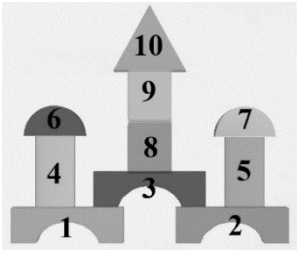

Fig. 5 The index of each piece in the castle game

2.3.3. Objective measurement method

Existing literature lacks standardized methods for objective measurements in CVE-based collaborative games. Measurements of collaboration and communication in CVE systems may be task-dependent.

Therefore, measurement methods used for existing CVE systems with social activities cannot be directly used in our CVE system with the collaborative puzzle games. One of the important goals of this chapter is to understand the collaboration and verbal-communication skills of children with ASD when they play collaborative games with their TD peers in CoMove. In order to understand the collaborative interactions and verbal-communication in the system, three kinds of data (game information data, human behaviors data and audio data) were recorded. The real-time recorded game information data included the start time, the end time and the success frequency of a game. The human behaviors data included information about dragging a piece, rotating a piece, and releasing a piece. The verbal-communication of each user in the environment was recorded as the audio data. These recorded data were analyzed offline based on several selected performance measures and communication measures.

Table 4 All the performance- and communication-related measures

	Measure name	Measure description
1	Success frequency	How many times an individual succeeded in moving pieces in game(s)
2	Collaborative movement ratio	The ratio of the time duration of a piece being moved by two individuals simultaneously to the time duration of an individual dragged the piece
3	Frequency of words	How many words per minute an individual spoke
4	Frequency of asking question	How often (the number of the utterance per minute) an individual asked task related questions
5	Frequency of information sharing-response	How often an individual responded to task related information
6	Frequency of information sharing-spontaneous	How often an individual initiated a task related information
7	Frequency of social reinforcement-positive	How often an individual used positive social reinforcement, such as “good job”.
8	Frequency of social reinforcement-negative	How often an individual used negative social feedback, such as “stupid”.
9	Frequency of directives	How often an individual directed other individual to tack action.
10	Frequency of social-oriented utterance	How often an individual used social-oriented utterance, such as “what is your name?”

The performance measures, as shown in Table 4, were chosen such that they could directly reflect the effects of the collaborative puzzle games and the collaboration strategies on users’ collaborative behaviors within the system. For example, the measure-collaborative movement ratio-was defined as the ratio of collaborative movement time (i.e., how long two users move puzzle pieces together) to total time

(i.e., how long an individual moves puzzle piece). The values of these performance measures were automatically computed offline from the recorded game information and human behavior data.

The communication measures, also shown in Table 4, were selected based on both previous studies in peer-based interactions (Teasley 1995) and the specific tasks in our environment. Previous studies have utilized several communication measures, including the number of spoken words (Teasley 1995), asking questions and answering questions (Van Boxtel et al. 2000), information sharing (Curtis and Lawson 2001), positive-reinforcement and negative-reinforcement (Mitchell et al. 2013), and directive utterances (Caballé et al. 2011), to understand both ASD and TD children in peer-based interactions. Based on these studies, we defined a corresponding seven communication measures to understand users' conversations about the game play in CoMove. Although the majority of conversations in multi-user interactions are task-oriented, non-task-oriented conversations (or social-oriented conversations) are also used by the users during interactions (Charlop-Christy et al. 2002). The non-task-oriented conversations were recorded using a *frequency of social-oriented utterance* measure. In order to index important aspects of communication in CoMove, we analyzed recorded audio data based on these measures by the following steps: i) transcribe the recorded audio data using the DragonNaturallySpeaking software (www.nuance.com), ii) correct the transcription by two native speakers of English, and iii) classify the corrected transcription into these communication measures by a human coder, who was blind with respect to the tasks and the participants.

In order to mitigate coding variability, our team of clinical psychologists and engineers collaborated to develop a rule-based coding protocol with structured instructions on how to classify these communication measures in a consistent manner. We provide two examples of predefined rules here. One, if the utterance starts with 'what', 'which', 'do', 'is', and 'are', it is classified as a question-asking utterance. Two, if an utterance provides color, position, rotation, direction, and puzzle piece information, it is an information sharing utterance. In addition, if the information sharing utterance has the same information as its preceding utterance, it is an information sharing-response utterance. Otherwise, it is an information sharing-spontaneous utterance. Please note that the human coder was trained and directly supervised by

two licensed clinical psychologists, who specialize in autism spectrum disorder intervention. The coder had been trained for behavior coding across a series of other related works(C. Liu et al. 2009; Zhang et al. 2017). By recording and analyzing these data from both performance and communication measures, this chapter provides a way to objectively measure collaboration and communication children with ASD within CoMove.

2.4. Feasibility Study

2.4.1. Subjects

A total of 28 children, 7 age- and gender-matched ASD/TD pairs and 7 age- and gender-matched TD/TD pairs (age range: 7 – 17 years) were recruited to participate. The rationale for the above grouping are as follows: i) the ultimate goal of the study is to understand and enhance the collaborative interaction and communication skills of children with ASD with their TD peers; and ii) we also wanted to explore how two TD children interact under the same conditions so that we can identify meaningful differences between TD and ASD interactions. All the children with ASD had a clinical diagnosis of ASD from a licensed clinical psychologist, an IQ higher than 70, and the ability to use phrased speech as determined by a trained therapist.

To assess current levels of ASD symptoms across groups, the Social Responsiveness Scale, second edition (SRS-2) (Constantino and Gruber 2002) and Social Communication Questionnaire (SCQ) (Rutter et al. 2003) were completed by parents of participants in both groups. These scales are efficient quantitative measures of interpersonal behavior, communication, and repetitive/stereotypic behavior characteristic of ASD. SRS-2 is an objective measure of symptoms associated with ASD. A total T-score of 76 or higher is considered strongly associated with clinical diagnosis of ASD. T-scores of 66 through 75 are interpreted as moderate deficiencies in reciprocal social behavior, whereas a T-score of 60 to 65 is interpreted as a mild range. Total scores in the range of 59T and below are generally not associated with clinically significant ASD. SCQ is a parent-reported screening measure that taps the symptomatology associated with ASD (Rutter et al. 2003). The SCQ has two versions, Lifetime and Current. A Lifetime

Total Score above 15 suggests that the individual is likely to have ASD. The Current Total Score is used to measure the individual's behaviors during the most recent 3-month period.

The characteristics of the 28 participants are shown in Table 5. TD1 represents the TD children in the ASD/TD pairs; while TD2 represents the TD children in the TD/TD pairs. These participants did not know each other before their experiments. They could not see each other but could talk with each other during the experiments through audio chat. The experiments were approved by the Vanderbilt University Institutional Review Board (IRB).

Table 5 Characteristics of The Participants

	Age Mean(SD)	Gender Female/male	SRS-2 total raw score Mean(SD)	SCQ current total score Mean(SD)
ASD	13.71(2.70)	1/6	107(22.35)	19(9.40)
TD1	13.89(3.14)	1/6	13.71 (16.06)	1.29(1.38)
TD2	10.59(2.00)	2/12	18.14 (16.60)	2.14 (3.53)

2.4.2. Tasks and protocol

Each pair of participants took part in a 50-minute long session where the participants sat in two different rooms in the same building. Two Dell desktop computers T3610 (E5-220 V3 CPU and 8GB RAM) were used as two nodes of CoMove. The connection of two nodes in these rooms was created via a Local Area Network (LAN). The experimental session included a pre-test, followed by a game playing session, and then a post-test. At the very beginning of an experiment, participants were instructed that they would be playing three different kinds of games in CoMove. However, they did not receive detailed instructions on the objective of each game. The participants were encouraged to communicate with each other in order to find out how to play each game. Feedback was also provided when necessary, as discussed in Section 2.3.2, in order to help participants play these games. After the introduction, the participants completed a pre-test that consisted of three baseline games (one castle game and two tangram games). The castle game is a turn-taking game requiring information sharing; while the two tangram games are enforced-collaboration games, T6 and T7, which are defined in Table 1. The target shapes of T6 and T7 in the pre-test were different from those in the game playing session. Eleven tangram games were included in the game playing session, which lasted approximately 30 minutes. After the game

playing session, the three baseline games were presented again to serve as the post-test. The games and the order of games during the experiment are shown in Table 6. During the experiment, we recorded both video and audio of the participants and their computer screens. The participants were informed that both audio and video recordings of the experiment would be made.

Table 6 The games and their order during one experiment

Pre-test	Castle game, T6, T7
Game playing	T1, T2, T3, T4, T5, T6, T7, T7, T6, T5, T4
Post-test	Castle game, T6, T7

2.5. Results and Discussions

2.5.1. System performance

Overall, CoMove worked as designed. All participants completed their experiments with a zero-dropout rate. The system mostly ran at 60 frames per second and had an average network delay of 1 millisecond or less when used through a Local Area Network (LAN). The system could successfully log the performance and audio data of the participants as well as game information data via the data logging component. The audio data were recorded for each game and each individual. A total of 467 audio files were recorded from 28 participants (17 audio files for each participant). One audio file was recorded incorrectly for unknown reasons. For the incorrectly recorded audio file, we manually extracted the corresponding audio data from the recorded video of that experiment. These results support the system feasibility. Specifically, the system is able to support communication and collaboration between children with ASD and their TD peers, as well as to record related data for meaningful measurements.

In order to index important aspects of interactions in CoMove, we analyzed changes of the predefined performance and communication measures (these measures are discussed in Section 2.3.3) from the pre-test to the post-test for each group. Specifically, we statistically compared the results of pre- and post-tests regarding these predefined performance and communication measures for all participants. The Wilcoxon Signed-rank test (Gibbons and Chakraborti 2011) was used for the statistical analysis with 0.05 as the alpha level. Given the limited power corresponding to this small sample size and conservative non-

parametric approach, we also examined effect sizes using Spearman's rank correlation (0.1 is small effect, 0.3 is medium effect, and 0.5 is large effect) (Cohen 1988). Although analyzing the differences between children with ASD and their TD peers in CVEs may be important, this chapter focuses on understanding changes of collaboration and verbal-communication skills within the system. Therefore, we compared the changes of each predefined measure from the pre-test to the post-test for each group.

2.5.2. Feasibility study results

The statistical analysis results for each subject group from the pre-test to the post-test across all performance measures are shown in Table 7 with the ρ columns showing all effect sizes. Overall, participants in all groups (i.e., children with ASD, TD1 children, and TD2 children) demonstrated statistically significant improvements on some performance measures from pre- to post-test. Wilcoxon Signed-rank test indicated a significantly higher collaborative movement ratio in the post-test relative to the pre-test in tangram games for children with ASD ($p<.05$, $\rho=.30$) and TD1 ($p<.05$, $\rho=.86$). ASD and TD1 also had a significantly higher success frequency in tangram games (two participants of a pair shared the same value regarding this frequency) in the post-test compared to the pre-test ($p<.05$, $\rho=.18$). In addition, Wilcoxon Signed-rank test indicated a significantly increased success frequency in the castle game regarding TD1 ($p<.05$, $\rho=.87$) and TD2 ($p<.001$, $\rho=.76$).

Table 8 summarizes the statistical differences from the pre-test to the post-test across all communication measures for all participants, while the ρ columns show all effect size results. Overall, all participants had changes in some communication measures from the pre-test to the post-test, although not all of these changes were statistically significant. The word frequency of children with ASD in castle game in the post-test is higher that the frequency in the pre-test, but not at a statistically significant level ($p=.25$, $\rho=.60$). TD1 had a significantly higher word frequency ($p<.05$, $\rho=.78$) in the castle game in the post-test compared to the pre-test. For children with ASD, the frequencies of asking questions and information sharing-spontaneous are higher in the post-test compared to the pre-test; however these

differences did not achieve statistical significance. TD1 children had a significantly higher frequency of information sharing-response in the castle game ($p<.05$, ρ =.29).

In summary, children with ASD, TD1 children, and TD2 children demonstrated improvements regarding some important collaboration and verbal-communication measures from the pre-test to the post-test, although not all improvements were statistically significant. In the next sub-section 2.5.3, we further discuss the results of children with ASD.

Table 7 Performance results from pre-tests to post-tests

		Castle game			Tangram games		
	Measure	Pre-test	Post-test	ρ	Pre-test	Post-test	ρ
ASD	1	2	5	0.81	7	14	0.18*
N=7	2	-	-	-	0.11	0.22	0.30*
TD1	1	3	4	0.87*	7	14	0.18*
N=7	2	-	-	-	0.11	0.12	0.86*
TD2	1	1	3.5	0.76**	8	12	0.30
N=14	2	-	-	-	0.13	0.09	0.22

*indicates $p<.05$ and ** indicates $p<.001$

The statistical analysis results for each subject group from the pre-test to the post-test across all performance measures are shown in Table 7 with the ρ columns showing all effect sizes. Overall, participants in all groups (i.e., children with ASD, TD1 children, and TD2 children) demonstrated statistically significant improvements on some performance measures from pre- to post-test. Wilcoxon Signed-rank test indicated a significantly higher collaborative movement ratio in the post-test relative to the pre-test in tangram games for children with ASD ($p<.05$, ρ=.30) and TD1 ($p<.05$, ρ=.86). ASD and TD1 also had a significantly higher success frequency in tangram games (two participants of a pair shared the same value regarding this frequency) in the post-test compared to the pre-test ($p<.05$, ρ=.18). In addition, Wilcoxon Signed-rank test indicated a significantly increased success frequency in the castle game regarding TD1 ($p<.05$, ρ=.87) and TD2 ($p<.001$, ρ=.76).

Table 8 summarizes the statistical differences from the pre-test to the post-test across all communication measures for all participants, while the ρ columns show all effect size results. Overall, all participants had changes in some communication measures from the pre-test to the post-test, although not

all of these changes were statistically significant. The word frequency of children with ASD in castle game in the post-test is higher that the frequency in the pre-test, but not at a statistically significant level (p=.25, ρ=.60). TD1 had a significantly higher word frequency (p<.05, ρ=.78) in the castle game in the post-test compared to the pre-test. For children with ASD, the frequencies of asking questions and information sharing-spontaneous are higher in the post-test compared to the pre-test; however these differences did not achieve statistical significance. TD1 children had a significantly higher frequency of information sharing-response in the castle game (p<.05, ρ =.29).

In summary, children with ASD, TD1 children, and TD2 children demonstrated improvements regarding some important collaboration and verbal-communication measures from the pre-test to the post-test, although not all improvements were statistically significant. In the next sub-section 2.5.3, we further discuss the results of children with ASD.

Table 8 The Communication Measures Results

		Castle game			Tangram games		
	Measure	Pre-test	Post-test	Rho	Pre-test	Post-test	Rho
ASD N=7	3	0.79	0.910	0.60	0.90	0.88	0.68
	4	0.02	0.014	0.84	8e-3	0.02	0.67
	5	0.05	0.069	0.46	0.05	0.04	0.89
	6	0.02	0.025	0.78	0.02	0.04	0.54
	7	0	0	0.45	0	0	0.15
	8	0	0	--	0	0	1
	9	0.01	0.014	0.52	0.01	0.02	0.59
	10	0	0	-0.17	0	0	--
TD1 N=7	3	0.73	1.038	0.78*	0.90	0.92	0.86
	4	0.03	0.035	0.68	0.01	0.01	0.86
	5	0.05	0.097	0.29*	0.05	0.03	0.29
	6	0.01	0.023	0.78	0.02	0.03	0.68
	7	0	0	--	2e-3	0	0.45
	8	0	0	--	0	0	--
	9	6e-3	0	0.61	8e-3	0.03	0.68
	10	0	0	--	0	0	--
TD2 N=14	3	0.74	0.863	0.81	0.75	0.58	0.58*
	4	0.01	0.020	0.40	0.02	0.01	0.72
	5	0.05	0.069	0.81	0.05	0.04	0.69
	6	0.01	0.018	0.65	0.01	0.03	0.67
	7	0	0	-0.16	6e-3	0	0.46*
	8	0	0	--	0	0	0.18
	9	6e-3	8e-3	0.14	0.02	0.02	0.63
	10	0	0	0.83*	0	0	0.07*

*indicates P<.05

2.5.3. Discussions

In spite of its small sample size and lack of an active comparison condition, results from this pilot study indicated that children with ASD showed improvements on some collaborative performance measures when using our system. In CoMove, children with ASD had a significantly increased success frequency in tangram games in the post-test compared to the pre-test. These results are consistent with previous work by Bauminger-Zviely et al. (Bauminger-Zviely et al. 2013), who had reported that collaborative games could improve collaborative performance of children with ASD. Battocchi et al. (Wilson and Russell 2007) mentioned that enforced collaboration games, which required simultaneous activities, could be used in collaboration training of children with ASD. We observed a statistically significant increase in the collaborative movement ratio for children with ASD. These results support the potential usability of the collaboration games to promote collaboration in the ASD population in the future.

In both castle and tangram games, children with ASD asked more questions in the post-test compared to the pre-test, even though the difference was not statistically significant. We noticed that in the pre-test of tangram games, the question asking frequency of children with ASD, $Mdn^3 = 8e-3$, was lower than the frequency of TD1 children, $Mdn = .01$, and TD2 children, $Mdn = .02$. This is consistent with Schmidt et al.'s findings (M Schmidt et al. 2011), which showed that children with ASD had a lower frequency of initiations, including question asking. However, in the post-test of our CVE puzzle games, the question asking frequency of children with ASD,$Mdn = .012$, was comparable to the frequency of TD1 children, $Mdn = .012$, and TD2 children, $Mdn = .011$. This result may be in line with Owen-Schryver's findings, which suggested that children with ASD could make more initiations after interacting with their TD peers in peer-mediated interactions (Owen-DeSchryver et al. 2008). However,

[3] Mdn is the median value. We show the median value because of the small sample size and non-normal distribution of the data.

these differences did not achieve statistical significance. The effects of CoMove on verbal-communication skills of children with ASD need to be further investigated in the future.

We also found that in both castle and tangram games, the success frequency of children with ASD was comparable to the frequency of TD children. The potential reason is that all participants with ASD had average IQ and phrase speech. Please note that this is not true for all children with ASD, some of whom have intellectual disability or severe language impairment. Therefore, these results should not be considered representative of how all children with ASD would perform.

Based on the above results and discussion, we cautiously conjecture that CoMove has the potential to index important aspects of interactions in the system. The effects of CoMove on collaborative and verbal-communication skills of children with ASD need to be further investigated in the future using a long-term, multi-session skill transfer study with more subjects. We believe that the results of this preliminary study warrant further exploration on the *potential impacts of CoMove on participants' skills in real life in a clinical study of generalization.*

2.6. Conclusions, Limitation, and Future Work

This chapter presents the design and development of a novel CVE system, CoMove, and the results of a preliminary feasibility study with 28 participants (7 ASD/TD pairs and 7 TD/TD pairs) using CoMove. A CVE system allows peer-based interaction in a shared and controlled virtual environment. The CVE system in this chapter has two distinguishing characteristics. First, it applies collaborative puzzle games and collaboration strategies to promote important collaborative behaviors, i.e., information sharing, sequential interactions, and simultaneous interactions. Second, it provides a potential way to objectively measure important aspects of collaboration and communication of these children when they play collaborative puzzle games in the CVE.

A total of 7 ASD/TD pairs and 7 TD/TD pairs were involved in a preliminary feasibility study to initially test system functionality, and determine its capacity for indexing important aspects of within-system interactions. Regarding functionality, all participants completed their experiments with a zero-

dropout rate. Additionally, all system software and hardware worked as designed, capturing necessary data with minimal loss. We also observed and measured changes in participant collaborative behaviors. Specifically, we observed statistically significant changes on some important performance measures in children with ASD: a significantly increased success frequency and a significantly increased collaborative movement ratio in tangram games. We also observed changes (although not statistically significant) in communication measures in the children with ASD, such as an increased frequency of question-asking in tangram games. These results support that our system has the potential to index important aspects of interactions in the system.

While the present work is promising, several limitations exist in the current work. We designed CoMove with the ultimate goals of measuring and potentially enhancing collaborative interactions and verbal-communication of children with ASD. In this chapter, we tested the system feasibility and its potential to measure important aspects of within-system interactions using a preliminary study. In the next step, we will evaluate how the within-system interactions correlate with and potentially impact participants' skills in real life in a clinical study of generalization. This future work will include long-term, multi-session experiments and many more participants. These participants will play real-world collaborative tasks as baseline tasks before and after interacting within CoMove. Their behavior changes in the baseline tasks will be measured using multiple measurement methods, such as questionnaires, observation, and evaluation tools (Gress et al. 2010), and used to indicate the impact of CoMove.

Second, only one coder coded the participants' communication behaviors in this study. In the current protocol, we were forced to rely on a single coder due to resource limitations in the laboratory of our behavioral collaborators in this preliminary work. In order for the coder to code communication behaviors in a consistent manner, our team of clinical psychologists and engineers collaborated to create predefined rules, as discussed in Section 2.3.4. However, the reliability of the coding results still needs to be evaluated using multiple coders and establishing high inter-rater reliability. We plan to use multiple coders in future work replicating and extending these finding.

Third, finding ways to provide and encourage face-to-face communication is a challenge for all CVEs. CoMove, at present, can only support audio chat, which simplified the data analysis with emphasis on the verbal communication in this preliminary study. Other non-verbal communication mode such as eye contact and gestures will be introduced in the future via a video chat, an eye tracker, and a gesture recognition method.

Fourth, CoMove was tested in a Local Area Network (LAN). In order to broaden its applicability, the system will need to be tested in a global area network in the future. Asymmetric latencies are issues associated with the server-client architecture. While the asymmetric latencies between the server node and the client node were small in the LAN, the asymmetric latencies need to be addressed in order for CoMove to be used in a global area network.

Finally, the current system included performance-based feedback. In the future, other kinds of feedback, such as communication-based feedback, will be included to foster more social-oriented communication. Despite the above-mentioned limitations, we believe that the present work, which offers a potential way to address current challenges of CVEs for ASD intervention and provides important preliminary insights in CVE and collaborative games-based intervention, makes a compelling case in this research area.

2.7. References

Anagnostou, E., Jones, N., Huerta, M., Halladay, A. K., Wang, P., Scahill, L., et al. (2015). Measuring social communication behaviors as a treatment endpoint in individuals with autism spectrum disorder. *Autism, 19*(5), 622-636.

Battocchi, A., Ben-Sasson, A., Esposito, G., Gal, E., Pianesi, F., Tomasini, D., et al. (2010). Collaborative puzzle game: a tabletop interface for fostering collaborative skills in children with autism spectrum disorders. *Journal of Assistive Technologies, 4*(1), 4-13.

Battocchi, A., Pianesi, F., Tomasini, D., Zancanaro, M., Esposito, G., Venuti, P., 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 (pp. 197-204): ACM

Bauminger-Zviely, N., Eden, S., Zancanaro, M., Weiss, P. L., & Gal, E. (2013). Increasing social engagement in children with high-functioning autism spectrum disorder using collaborative technologies in the school environment. *Autism, 17*(3), 317-339.

Bauminger, N., Goren-Bar, D., Gal, E., Weiss, P. L., Kupersmitt, J., Pianesi, F., et al. Enhancing social communication in high-functioning children with autism through a co-located interface. In *Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on, 2007* (pp. 18-21): IEEE

Ben-Sasson, A., Lamash, L., & Gal, E. (2013). To enforce or not to enforce? The use of collaborative interfaces to promote social skills in children with high functioning autism spectrum disorder. *Autism, 17*(5), 608-622.

Benford, S., Greenhalgh, C., Rodden, T., & Pycock, J. (2001). Collaborative virtual environments. *Communications of the ACM, 44*(7), 79-85.

Bernard-Opitz, V., Sriram, N., & Nakhoda-Sapuan, S. (2001). Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. *Journal of autism and developmental disorders, 31*(4), 377-384.

Bowers, J., Pycock, J., & O'brien, J. Talk and embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1996* (pp. 58-65): ACM

Boyd, L. E., Rangel, A., Tomimbang, H., Conejo-Toledo, A., Patel, K., Tentori, M., et al. SayWAT: Augmenting Face-to-Face Conversations for Adults with Autism. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2016* (pp. 4872-4883): ACM

Boyd, L. E., Ringland, K. E., Haimson, O. L., Fernandez, H., Bistarkey, M., & Hayes, G. R. (2015). Evaluating a collaborative iPad game's impact on social relationships for children with autism spectrum disorder. *ACM Transactions on Accessible Computing (TACCESS), 7*(1), 3.

Burke, M., Kraut, R., & Williams, D. Social use of computer-mediated communication by adults on the autism spectrum. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work, 2010* (pp. 425-434): ACM

Caballé, S., Daradoumis, T., Xhafa, F., & Juan, A. (2011). Providing effective feedback, monitoring and evaluation to on-line collaborative learning discussions. *Computers in Human Behavior, 27*(4), 1372-1381.

Caltagirone, S., Keys, M., Schlief, B., & Willshire, M. J. (2002). Architecture for a massively multiplayer online role playing game engine. *Journal of Computing Sciences in Colleges, 18*(2), 105-116.

Cappelletti, A., Gelmini, G., Pianesi, F., Rossi, F., & Zancanaro, M. Enforcing cooperative storytelling: First studies. In *Advanced Learning Technologies, 2004. Proceedings. IEEE International Conference on, 2004* (pp. 281-285): IEEE

Charlop‐Christy, M. H., Carpenter, M., Le, L., LeBlanc, L. A., & Kellet, K. (2002). Using the picture exchange communication system (PECS) with children with autism: Assessment of PECS acquisition, speech, social‐communicative behavior, and problem behavior. *Journal of applied behavior analysis, 35*(3), 213-231.

Cheng, Y., Chiang, H.-C., Ye, J., & Cheng, L.-h. (2010). Enhancing empathy instruction using a collaborative virtual learning environment for children with autistic spectrum conditions. *Computers & Education, 55*(4), 1449-1458.

Cheng, Y., & Ye, J. (2010). Exploring the social competence of students with autism spectrum conditions in a collaborative virtual learning environment–The pilot study. *Computers & Education, 54*(4), 1068-1077.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences Lawrence Earlbaum Associates. *Hillsdale, NJ*, 20-26.

Constantino, J. N., & Gruber, C. P. (2002). The social responsiveness scale. *Los Angeles: Western Psychological Services*.

Curtis, D. D., & Lawson, M. J. (2001). Exploring collaborative online learning. *Journal of Asynchronous learning networks, 5*(1), 21-34.

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. Heerlen, Open Universiteit Nederland.

Dix, A. (2009). *Human-computer interaction*: Springer.

Fan, M., Antle, A. N., Neustaedter, C., & Wise, A. F. Exploring how a co-dependent tangible tool design supports collaboration in a tabletop activity. In *Proceedings of the 18th International Conference on Supporting Group Work, 2014* (pp. 81-90): ACM

Fleury, C., Duval, T., Gouranton, V., & Arnaldi, B. Architectures and mechanisms to maintain efficiently consistency in collaborative virtual environments. In *SEARIS 2010 (IEEE VR 2010 Workshop on Software Engineering and Architectures for Realtime Interactive Systems), 2010*

Gal, E., Bauminger, N., Goren-Bar, D., Pianesi, F., Stock, O., Zancanaro, M., et al. (2009). Enhancing social communication of children with high-functioning autism through a co-located interface. *Ai & Society, 24*(1), 75-84.

Gal, E., Goren-Bar, D., Gazit, E., Bauminger, N., Cappelletti, A., Pianesi, F., et al. Enhancing social communication through story-telling among high-functioning children with autism. In *International Conference on Intelligent Technologies for Interactive Entertainment, 2005a* (pp. 320-323): Springer

Gal, E., Goren-Bar, D., Gazit, E., Bauminger, N., Cappelletti, A., Pianesi, F., et al. (2005b). Enhancing social communication through story-telling among high-functioning children with autism. In *Intelligent Technologies for Interactive Entertainment* (pp. 320-323): Springer.

Gautier, L., & Diot, C. Design and evaluation of mimaze a multi-player game on the internet. In *Multimedia Computing and Systems, 1998. Proceedings. IEEE International Conference on, 1998* (pp. 233-236): IEEE

Gibbons, J. D., & Chakraborti, S. (2011). *Nonparametric statistical inference*: Springer.

Gress, C. L., Fior, M., Hadwin, A. F., & Winne, P. H. (2010). Measurement and assessment in computer-supported collaborative learning. *Computers in Human Behavior, 26*(5), 806-814.

Hourcade, J. P., Bullock-Rest, N. E., & Hansen, T. E. (2012). Multitouch tablet applications and activities to enhance the social skills of children with autism spectrum disorders. *Personal and Ubiquitous Computing, 16*(2), 157-168.

Johnson, D. W., & Johnson, R. T. (1996). Cooperation and the use of technology. *Handbook of research for educational communications and technology: A project of the Association for Educational*

Communications and Technology, 1017-1044.

Laffey, J., Schmidt, M., Galyen, K., & Stichter, J. (2012). Smart 3D collaborative virtual learning environments: A preliminary framework. *Journal of Ambient Intelligence and Smart Environments, 4*(1), 49-66.

Leman, P. J. (2015). How do groups work? Age differences in performance and the social outcomes of peer collaboration. *Cognitive science, 39*(4), 804-820.

Liu, C., Agrawal, P., Sarkar, N., & Chen, S. (2009). Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *International Journal of Human-Computer Interaction, 25*(6), 506-529.

Liu, E. S., & Theodoropoulos, G. K. (2014). Interest management for distributed virtual environments: A survey. *ACM Computing Surveys (CSUR), 46*(4), 51.

Livingstone, D., Kemp, J., & Edgar, E. (2008). From Multi-User Virtual Environment to 3D Virtual Learning Environment. *ALT-J: Research in Learning Technology, 16*(3), 139-150.

Macedonia, M. R., & Zyda, M. J. (1997). A taxonomy for networked virtual environments. *IEEE multimedia*(1), 48-56.

McPartland, J. C., Reichow, B., & Volkmar, F. R. (2012). Sensitivity and specificity of proposed DSM-5 diagnostic criteria for autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(4), 368-383.

Millen, L., Cobb, S., Patel, H., & Glover, T. Collaborative virtual environment for conducting design sessions with students with autism spectrum conditions. In *Proc. 9th Intl Conf. on Disability, Virtual Reality and Assoc. Technologies, 2012* (pp. 269-278)

Millen, L., Hawkins, T., Cobb, S., Zancanaro, M., Glover, T., Weiss, P. L., et al. Collaborative technologies for children with autism. In *Proceedings of the 10th International Conference on Interaction Design and Children, 2011* (pp. 246-249): ACM

Mitchell, C. M., Ha, E. Y., Boyer, K. E., & Lester, J. C. (2013). Learner characteristics and dialogue: recognising effective and student-adaptive tutorial strategies. *International Journal of Learning Technology 25, 8*(4), 382-403.

Montoya, M. M., Massey, A. P., & Lockwood, N. S. (2011). 3D collaborative virtual environments: exploring the link between collaborative behaviors and team performance. *Decision Sciences, 42*(2), 451-476.

Moore, D., Cheng, Y., McGrath, P., & Powell, N. J. (2005). Collaborative virtual environment technology for people with autism. *Focus on Autism and Other Developmental Disabilities, 20*(4), 231-243.

Nakano, M., Sato, S., Komatani, K., Matsuyama, K., Funakoshi, K., & Okuno, H. G. A two-stage domain selection framework for extensible multi-domain spoken dialogue systems. In *Proceedings of the SIGDIAL 2011 Conference, 2011* (pp. 18-29): Association for Computational Linguistics

Natkin, S., & Yan, C. User model in multiplayer mixed reality entertainment applications. In *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology, 2006* (pp. 85): ACM

Newbutt, N. Representations of Self in Classroom Virtual Worlds: a case-study of pupils on the autism spectrum.

Noor, H. A. M., Shahbodin, F., & Pee, N. C. (2012). Serious game for autism children: Review of literature. *World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, 6*(4), 554-559.

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.

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

Reynolds, S., Bendixen, R. M., Lawrence, T., & Lane, S. J. (2011). A pilot study examining activity participation, sensory responsiveness, and competence in children with high functioning autism spectrum disorder. *Journal of autism and developmental disorders, 41*(11), 1496-1506.

Ringland, K. E., Wolf, C. T., Faucett, H., Dombrowski, L., & Hayes, G. R. Will I always be not social?: Re-Conceptualizing Sociality in the Context of a Minecraft Community for Autism. In *Proceedings of*

the 2016 CHI Conference on Human Factors in Computing Systems, 2016 (pp. 1256-1269): ACM

Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *The Journal of the Learning Sciences, 14*(2), 201-241.

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*: Western Psychological Services.

Schmidt, M., & Beck, D. Computational Thinking and Social Skills in Virtuoso: An Immersive, Digital Game-Based Learning Environment for Youth with Autism Spectrum Disorder. In *International Conference on Immersive Learning, 2016* (pp. 113-121): Springer

Schmidt, M., Laffey, J., & Stichter, J. Virtual social competence instruction for individuals with autism spectrum disorders: Beyond the single-user experience. In *Proceedings of CSCL, 2011* (pp. 816-820)

Schmidt, M., Laffey, J. M., Schmidt, C. T., Wang, X., & Stichter, J. (2012). Developing methods for understanding social behavior in a 3D virtual learning environment. *Computers in Human Behavior, 28*(2), 405-413.

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2009). The neural basis of multisensory integration in the midbrain: its organization and maturation. *Hearing research, 258*(1), 4-15.

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.

Suthers, D. D. Architectures for computer supported collaborative learning. In *Advanced Learning Technologies, 2001. Proceedings. IEEE International Conference on, 2001* (pp. 25-28): IEEE

Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology, 31*(2), 207.

Van Boxtel, C., Van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and instruction, 10*(4), 311-330.

Wallace, S., Parsons, S., & Bailey, A. (2015). Self-reported sense of presence and responses to social

stimuli by adolescents with ASD in a collaborative virtual reality environment. *Journal of Intellectual and Developmental Disability*.

Wang, X., Laffey, J., Xing, W., Ma, Y., & Stichter, J. (2016). Exploring embodied social presence of youth with Autism in 3D collaborative virtual learning environment: A case study. *Computers in Human Behavior, 55*, 310-321.

Weiss, M. J., & Harris, S. L. (2001). Teaching social skills to people with autism. *Behavior modification, 25*(5), 785-802.

Weiss, P. L., Gal, E., Zancanaro, M., Giusti, L., Cobb, S., Millen, L., et al. Usability of technology supported social competence training for children on the autism spectrum. In *Virtual Rehabilitation (ICVR), 2011 International Conference on, 2011* (pp. 1-8): IEEE

White, S. W., Keonig, K., & Scahill, L. (2007). Social skills development in children with autism spectrum disorders: A review of the intervention research. *Journal of autism and developmental disorders, 37*(10), 1858-1868.

Wilson, G. F., & Russell, C. A. (2007). Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. *Human factors: the journal of the human factors and ergonomics society, 49*(6), 1005-1018.

Zancanaro, M., Pianesi, F., Stock, O., Venuti, P., Cappelletti, A., Iandolo, G., et al. (2007). Children in the museum: an environment for collaborative storytelling. In *PEACH-Intelligent Interfaces for Museum Visits* (pp. 165-184): Springer.

Zhang, L., Wade, J., Bian, D., Fan, J., Swanson, A., Weitlauf, A., et al. (2017). Cognitive load measurement in a virtual reality-based driving system for autism intervention. *IEEE Transactions on Affective Computing, 8*(2), 176-189.

CHAPTER III. A COLLABORATIVE VIRTUAL ENVIRONMENT ON THE ANDROID PLATFOMR

3.1. Abstract

The previous chapter explored a Collaborative Virtual Environment (CVE) for children with Autism Spectrum Disorder (ASD) to interact with their Typically Developing (TD) peers from different locations using their own computers. Recently, there has been growing interest in mobile applications, which have the potential to increasingly engage children with ASD (Tanaka et al. 2010) by creating ubiquitous learning environments (Gravenhorst et al. 2015). In this chapter, we designed a Collaborative Virtual Environment (CVE) on the Android platform in order to investigate the collaborative behaviors and communication skills of children with ASD. The CVE on the Android platform 1) has widespread availability, and 2) allows flexible communication between people. This presented CVE on the Android platform allows two users in different locations to interact and communicate with each other while playing puzzle games on mobile devices. Multiple puzzle games with different interaction patterns were designed in the environment, including turn-taking, information sharing, and enforced collaboration. Audio and video chat were implemented in the environment in order for the geographically distributed players to talk with and see each other. The usability of the environment has been validated through a user study involving five pairs of subjects. Each pair included one child with ASD and one typically developing (TD) child. The results showed that the presented CVE environment may have the potential to improve players' collaborative behaviors and communication skills.

3.2. Introduction

In this chapter, we designed a CVE on the Android platform for ASD intervention with the goal to understand and ultimately promote the collaborative interactions and communications of children with ASD. Because the use of mobile devices is growing exponentially, mobile applications have the potential to increasingly engage children with ASD (Tanaka et al. 2010) by creating ubiquitous learning

environments (Gravenhorst et al. 2015). The studies using mobile applications for ASD intervention include emotion recognition (Leijdekkers et al. 2013), social interactions (Escobedo et al. 2012), and vocabulary learning (Husni 2013). However, these mobile applications are limited in interactions and communications between a user and a mobile device. Our CVE on the Android platform (Fig. 6) — supporting multiple players' interactions and communications in the shared collaborative environment— investigated the collaborative interactions and communications of children with ASD.

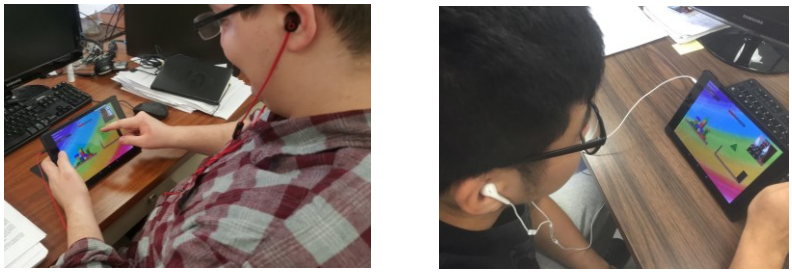

Fig. 6 Two players are using the CVE on the Android platform.

The goals of the current research were to: 1) evaluate the usability of the CVE on the Android platform; and 2) investigate the collaborative interactions and the communications of children with ASD when playing puzzle games with their TD peers. This novel environment supports following functionalities:

i) Interaction between two geographically distributed players via internet

ii) Audio and video communication

iii) Automatic performance and audio recording

3.3. Method

The CVE on the Android platform was designed with Unity3D (www.unity3d.com). The application, which can be accessed by players using Android mobile devices, allows two players to interact and communicate with each other remotely. A variety of puzzle games were developed that compelled interactions between two players, such as turn-taking, information sharing, and enforced collaboration. The environment supported video and audio communication between the two players. The performance status and dialogue of the players were recorded by the CVE system for offline analysis. In order to

support these functionalities, a software framework (Fig. 7) was developed with four modules: a game controller module, a network connection module, a communication module, and a data recording module. The game controller module implemented the application logic of the puzzle games. The audio/video chat between two geographically distributed players was supported through the communication module. The network connection module was responsible for transferring the data from game controller and communication modules via the internet. The data recording module recorded locally information related to how players played the game and how they communicated with each other. The functionality and the usability of this CVE on the Android platform has been evaluated by a small user study involving five pairs of players, each consisting of one child with ASD and one TD child.

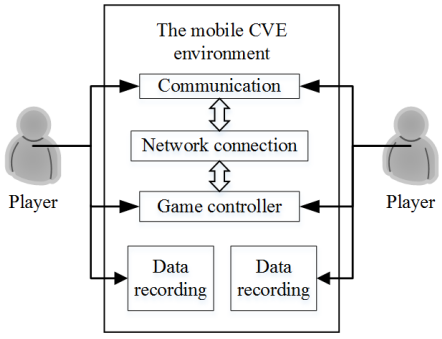

Fig. 7 The framework of our CVE on the Android platform.

3.3.1. System design

3.3.1.1. Game controller.

The logic of the CVE was implemented based on a hierarchical and concurrent Finite State Machine (FSM) model, shown in Fig. 8. The element of concurrency in the design made it possible for players to reside in different sub-states while still maintaining application synchronization. Take the 'Introduction' state for instance: player1 could stay in the 'Wait_P1' state after he/she finished reading the introduction, while player2 could still read the introduction in the 'Introduction_P2' state. After both players finished reading the introduction, their game states would be rejoined upon exiting the 'Introduction' state. In Fig. 8, M is the total number of puzzle games to be played in a session, which is customizable for different requirements. N is the total steps of one game, which can vary from game to game.

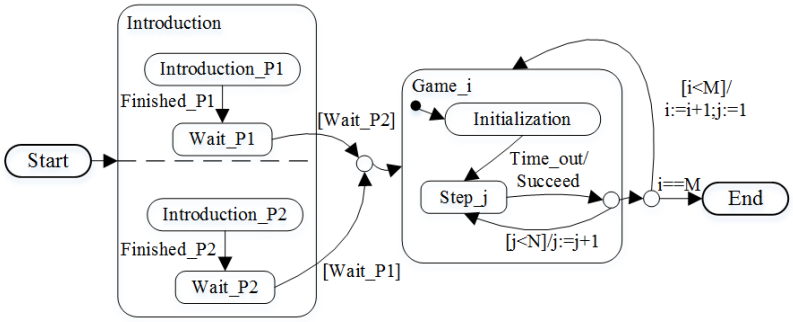

Fig. 8 The Finite State Machine model of the CVE on the Android platform.

The hierarchical state, 'Game_i', defined the logic of each puzzle game in the CVE. In order to study the different collaborative interactions of children with ASD, three types of puzzle games were designed in the CVE: turn-taking games, information sharing games, and enforced collaboration games. In the turn-taking games, players had alternating control of blocks movement. Information sharing games asked players to share color information in order to move the correct blocks. The enforced collaboration games aimed to impose simultaneous interactions of two players. Both players had to move the same block in the same direction at the same time in the enforced collaboration games. A considerable amount of communication between the players was required in the enforced collaboration games.

Table 9 Game Configuration Parameters

Game index	Color visibility	Block controllability	Rotatability	Interaction type
1	P1 and P2	P1 in step 1, 3, 5, 7; P2 in step 2, 4, 6	Auto	Turn-taking
2	P1 and P2	P1 in step 3, 4, 7; P2 in step 1, 2, 5, 6	Auto	Turn-taking
3	P1 and P2	P1 and P2 in all steps	Auto	Enforced collaboration
4	P1	P2 in all steps	P2	Information-sharing
5	P2	P1 in all steps	P1	Information-sharing
6	P1	P1 and P2 in all steps	P2	Enforced collaboration
7	P2	P1 and P2 in all steps	P1	Enforced collaboration
8	Half for P1 and Half for P2	P1 in step 1, 3, 5, 7, 9; P2 in step 2, 4, 6, 8, 10	Auto	Turn-taking

Each game was composed of multiple steps. The steps were generalized and modeled using the 'Step_j' state in Fig. 8. The implementation of the steps was different depending on the configuration of the following parameters: 1) color visibility (i.e., which players can see the color); 2) block controllability (i.e., which players can move the block); and 3) block rotatability (i.e., which players can rotate the

block). The values of these parameters determined the types of game interaction, shown in Table 9. Game 1 was a turn-taking game, in which two players, P1 and P2, were both able to see the color of blocks. At the first step, player P1 had control of all the blocks. At the next step, P2 had control of all the blocks. They would take turns dragging the block in game 1. The number of steps in each game was seven in all but the last game, which had ten steps.

3.3.1.2. Network connection.

A client-server network architecture was used for the CVE on the Android platform involving two players. The device of one player acted as a server, while the other player's device acted as a client. All the computationally-intensive tasks were implemented on the server side. This network architecture has been widely used for two-player mobile games and it is simple and sufficient (Gautier and Diot 1998). The network connection was created via Unity Master Server (UMS) (http://docs.unity3d.com/Manual/net-MasterServer.html), which allowed players to find each other at any time and at any location. The function of the UMS for network connection is shown in Fig. 9. Any player can initiate a game as a server, for example the 'Server A' or the 'Server B' in Fig. 9, by clicking the 'start game' button in the game. The device information (IP address and port number) were registered with the UMS. Other players can connect to one of the active servers, thus becoming a client. The client can send connection requests by clicking the 'connect' button in the game. After receiving this request, the UMS then returns a list of active servers to the client. The returned list contains all pertinent information required for the client to connect to the server. The client then selects one server from the list to create a connection with the selected server.

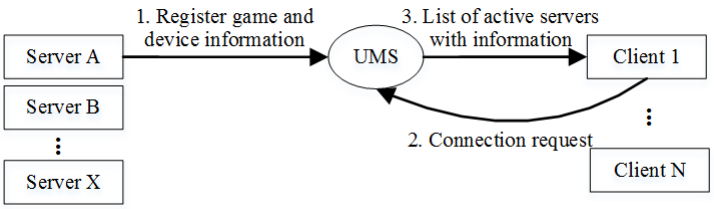

Fig. 9 Unity Master Server (UMS) role.

3.3.1.3. Communication module.

The communication module was designed to support the audio and video communication between the players. The video chat functionality allowed two players to see each other in real time, which was implemented following the procedure in Fig. 10. The image was captured by the mobile camera in ARGB32 format. The captured image was then encoded into a JPG file, which was amenable to network transfer because it had a small data size and was in serialized format. The Unity remote procedure call (http://docs.unity3d.com/Manual/net-RPCDetails.html) was used for the image data transferring. When the receiving mobile device obtained the transferred data, it was decoded into the RGB24 format and displayed using the Texture2D component (http://docs.unity3d.com/ScriptReference/Texture2D.html). The video was updated with a fixed frequency 12Hz.

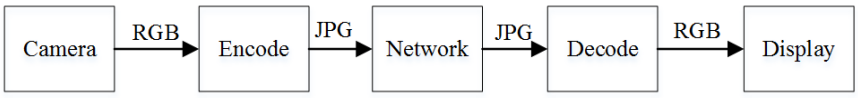

Fig. 10 The video chat procedure

The audio chat feature was used for players to talk with each other in real time. This feature was implemented with a procedure similar to the video chat. With both a speaker and microphone on the device, there was an audio feedback problem during the audio chat. Head phones were worn by all players to solve this audio feedback problem.

3.3.1.4. Data recording.

The data recording module records the performance and dialogue data of players in the CVE for the offline analysis. The recorded performance metrics included the success frequency (how many times they succeed per game), game duration, and collaborative movement duration, which were written to a file in real time. The audio data of each player were recorded locally with a frequency of 44.1 KHz. The recorded audio data were written to a file at the very end of all games, and were later transcribed and labeled with some predefined utterance types manually. The utterance types were defined referring to previous literature to evaluate the communication (Charlop-Christy et al. 2002; Nunes and Hanline 2007),

including the words spoken per minute, the frequency of question-asking, the frequency of response, the frequency of spontaneous information sharing, and the frequency social oriented utterance.

3.3.2. Experiment Setup

Five age- and gender-matched pairs of subjects were recruited for a preliminary evaluation of the CVE on the Android platform. Each pair was composed of one child with ASD and one TD child. All the children with ASD had a clinical diagnosis of ASD from a licensed clinical psychologist. The Social Responsiveness Scale, second edition (SRS-2) (Constantino and Gruber 2002) and Social Communication Questionnaire Lifetime Total Score (SCQ) (Rutter et al. 2003) were completed by for a parent of each child with ASD. The experiments were approved by the Vanderbilt University Institutional Review Board (IRB). The information of all the subjects are summarized in the Table 10.

Table 10 Subject Characteristics

	Age: Mean (Std)	Gender Female/male	SRS-2 total raw score: Mean (Std)	SCQ current total score: Mean (Std)
ASD	10.99 (3.69)	4/1	83.50 (24.96)	20.25 (10.50)
TD	10.81 (2.32)	4/1	9.80 (7.53)	0.80 (1.30)

During the experiment, subjects in a pair sat separately in two different rooms. The layout of the experiment rooms is shown in Fig. 11(left). One tablet, Nexus 9, and one set of headphones were provided for each subject for the experiment. A video camera recorded the subject and the device during the experiment. Each experiment lasted about 40 minutes. The experiment procedure is shown in Fig. 11(right). At the beginning, the devices were given to each subject. Then, subjects completed the pre-test, with one turn-taking game (game 8), and two enforced collaboration games (game 6 and game 7). During the core task, seven puzzle games, from game 1 to game 7, were presented in order. The post-test included the same games as the pre-test. After the post-test, subjects completed a survey regarding their experience with the system and with their partners.

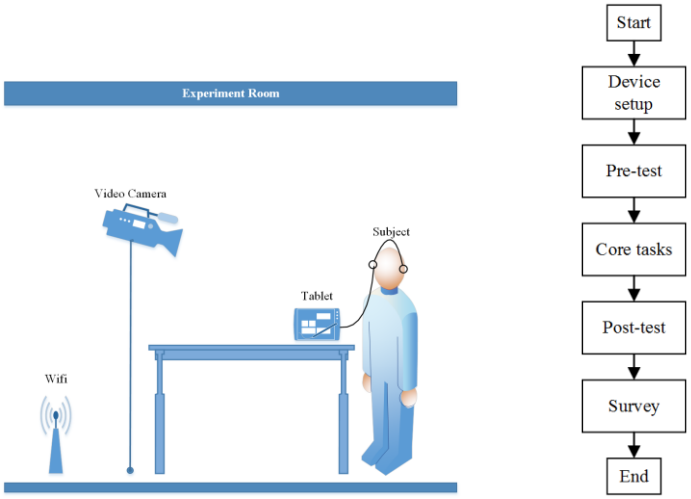

Fig. 11 The experiment room layout (left) and the experiment procedure (right).

3.4. Results

3.4.1. System performance

Five pairs completed the preliminary study. One pair had difficulty to finish the experiment and their data were excluded from analysis. The other four pairs completed the experiment. The network connection was lost during one pre-test. The automatically recorded performance and audio data of game 7 in this pre-test were also lost. However, we recovered the audio data by extracting the audio from the recorded video. The turn-taking game (game 8) and the enforced collaboration games (game 6 and game 7) were analyzed separately since different games required different interactions. Four pairs' data were used for the turn-taking game, the survey analysis, and the communication analysis of the enforced collaboration games. Only three pairs' data were used for the performance analysis in the enforced collaboration games. Even though the video chat was implemented in our CVE on the Android platform, it was disabled during the experiment because it caused data loss and lag. Instead, audio communication was used during the experiment.

3.5.2. Feasibility study results

All pairs showed an improved performance during the post-test compared to the pre-test. This meant all pairs succeeded more in both the turn-taking game (game 8), and the enforced collaboration games

(game 6 and game 7). In addition, all pairs required less time to finish both turn-taking game, and enforced collaboration games in the post-test. All pairs also demonstrated an increased collaborative movement duration, which was defined as the duration of time players moved blocks together. The increased collaborative duration may indicate improved collaboration between the players. The mean and standard deviation (Std) of all these changes are listed in Table 11. The first row shows that the success frequency increased by 3.25 seconds on average in game 8, which allowed a maximum of 10 successes, with a standard deviation 1.7.

Table 11 Performance Changes from Pre-test to Post-test

Game Index	Variable	Increased/ Decreased	Mean (Std)
8	Success frequency	Increased	3.25 (1.70)
	Time duration (in seconds)	Decreased	28.00 (18.78)
6 & 7	Success frequency	Increased	8.67 (3.06)
	Time duration (in seconds)	Decreased	304.67 (83.94)
	collaborative ratio (in seconds)	Increased	20.62 (8.55)

For the communication data, we found some differences between the children with ASD and their TD peers during the pre-test in the enforced collaboration games. In the pre-test, all the children with ASD spoke fewer words than their TD partners in the enforced collaboration games. They also asked fewer questions, but gave more responses, compared to their TD partners in the pre-test of the enforced collaboration games. The mean and standard deviation of the absolute value of these differences are listed in Table 12. However, these differences were not observed during the post-test. The changes from the pre-test to the post-test in terms of the communication variables were not found.

Table 12 Difference between Children with ASD and Their TD Partners during the Pre-test

	Words	Questions	Responses
Mean (Std)	39.75 (27.62)	8.00 (8.66)	3.00 (2.65)

The survey from the subjects reflected positive opinions for the environment. From the survey, all subjects enjoyed playing the games. All subjects perceived an improved individual performance in playing the games and increased ease in talking with their partners at the end of the experiments. Each question in the survey was scored on a 5-Likert scale. In terms of enjoying the game, number 1 indicated

no enjoyment at all, while number 5 meant very much enjoyment. In terms of their performance in the game and the communication with their partners, number 1 indicated performance/communication became much worse at the end of the experiment, while number 5 meant performance/communication became much better at the end of the experiment. The mean and standard deviation of the survey questions are shown in Table 13.

Table 13 Self-report Results

	Enjoyment	Increased communication ease	Improved individual performance
Mean(Std)	4.88 (0.35)	4.63 (0.74)	4.50 (0.53)

3.5. Conclusions and Future Works

This work discusses the design of a CVE on the Android platform for ASD intervention with the goal to improve the collaborative interaction and communications of children with ASD using a mobile application. The environment facilitates the interaction and communication of two players from different locations by playing multiple block games.

Five ASD/TD pairs participated in the preliminary study. The usability of the environment and its functionalities, including the two players' interactions, audio communications, and data recording, have been validated by the preliminary study. The result of the experiments may support the potential of the environment in improving the collaborative interactions and communications of children with ASD.

There were some limitations in the system design and user study in this work. Only a small sample size were included in our current study. More subjects will be involved for the experiments in the future. The current CVE was evaluated in a small local network, which will be extended to the global network in order for players from any location in the world to gain access. Future work will consist of a closed-loop system using targeted feedback based on the player's communication and collaborative performance.

3.6. References

Charlop‐Christy, M. H., Carpenter, M., Le, L., LeBlanc, L. A., & Kellet, K. (2002). Using the picture exchange communication system (PECS) with children with autism: Assessment of PECS acquisition,

speech, social‐communicative behavior, and problem behavior. *Journal of applied behavior analysis, 35*(3), 213-231.

Constantino, J. N., & Gruber, C. P. (2002). The social responsiveness scale. *Los Angeles: Western Psychological Services*.

Escobedo, L., Nguyen, D. H., Boyd, L., Hirano, S., Rangel, A., Garcia-Rosas, D., et al. MOSOCO: a mobile assistive tool to support children with autism practicing social skills in real-life situations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2012* (pp. 2589-2598): ACM

Gautier, L., & Diot, C. Design and evaluation of mimaze a multi-player game on the internet. In *Multimedia Computing and Systems, 1998. Proceedings. IEEE International Conference on, 1998* (pp. 233-236): IEEE

Gravenhorst, F., Muaremi, A., Bardram, J., Grünerbl, A., Mayora, O., Wurzer, G., et al. (2015). Mobile phones as medical devices in mental disorder treatment: an overview. *Personal and Ubiquitous Computing, 19*(2), 335-353.

Husni, E. (2013). Mobile Applications BIUTIS: Let's Study Vocabulary Learning as a Media for Children with Autism. *Procedia Technology, 11*, 1147-1155.

Leijdekkers, P., Gay, V., & Wong, F. CaptureMyEmotion: A mobile app to improve emotion learning for autistic children using sensors. In *Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on, 2013* (pp. 381-384): IEEE

Nunes, D., & Hanline, M. F. (2007). Enhancing the Alternative and Augmentative Communication Use of a Child with Autism through a Parent‐implemented Naturalistic Intervention. *International Journal of Disability, Development and Education, 54*(2), 177-197.

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*: Western Psychological Services.

Tanaka, J. W., Wolf, J. M., Klaiman, C., Koenig, K., Cockburn, J., Herlihy, L., et al. (2010). Using computerized games to teach face recognition skills to children with autism spectrum disorder: the Let's Face It! program. *Journal of Child Psychology and Psychiatry, 51*(8), 944-952.

CHAPTER IV. DESIGN OF AN INTELLIGENT AGENT FOR MEASUREMENTS IN A CVE

4.1. Abstract

In Chapter II and Chapter III, we designed and developed Collaborative Virtual Environments (CVEs) in order to encourage collaboration and communication between real-users. Although CVEs have the advantages to support flexible, safe and peer-based interactions, measuring the interactions in CVEs is challenging given the complex interactions and unrestricted conversations between the real-users. In this chapter, we have designed an intelligent agent that could communicate and play games with users in order to measure their communication and collaboration skills in a CVE. The intelligent agent was developed with a hybrid method, which combined a dialogue act classifier and a finite state machine. This hybrid method enabled the intelligent agent not only to communicate and play collaborative puzzle games with the users in the CVE but also to generate task-performance and verbal-communication features to measure their both communication and collaboration skills. A preliminary study with five children with ASD was conducted to test the intelligent agent. Results demonstrated the capacity of the intelligent agent to communicate and play games with children, as well as the potential to generate meaningful features to measure the skills.

4.2. Introduction

As discussed in Chapter II and Chapter III, a Collaborative Virtual Environment (CVE), which is a computer-based, distributed, virtual space for multiple-users to interact with one another and/or with the virtual items (Benford et al. 2001), preserves the advantages of traditional computer-based intervention systems but also facilitate real-time interactions between real users across distance. CVE technology offers a flexible alternative to conventional modalities of both in-vivo (e.g., social skill groups, peer-mediated programs) and technological intervention (e.g., confederate controlled virtual reality (VR), computerized skill programs) where multiple individuals can share and interact in a virtual space. In

particular, the characteristics of this environment are highly controllable and can be adapted and structured in ways that mimic aspects of real-world interactions. These characteristics can tangibly impact the very nature of the collaborative interaction itself.

Although CVEs provide a promising platform for realistic interactions between real users, CVE-based interventions lack reliable and easy-to-use methods to measure social communication within these systems. The majority of CVEs in this area measured users' behaviors within the systems based on self-report questionnaires or their task-performance. For example, Wallace et al. designed and developed a CVE-system to teach greeting behaviors to children with ASD in a virtual gallery (Wallace et al. 2015). They evaluated the system impacts using a self-report questionnaire, and found that children with ASD, compared to their Typically Developing (TD) peers, were less sensitive to negative greetings. Millen et al. applied CVEs to promote collaboration among children with ASD, and the results of a self-report questionnaire showed improved engagement of children with ASD in the CVEs (Millen et al. 2011). Cheng et al. designed a CVE-based virtual restaurant to understand empathy of children with ASD (Cheng et al. 2010). They found that these children could appropriately answer more empathic questions after the intervention. Although these methods could gather essential information for system evaluation, they could not be used to understand and analyze users' conversation, which is important aspects of user-to-user interactions in most CVE-based interventions.

In some instances, domain experts have been involved to observe and code not only task-performance but also verbal communication of users in CVEs using a human coding methodology. iSocial is a 3D-CVE aimed at understanding and improving social competency development of children with ASD (M Schmidt et al. 2011). In iSocial, children's social behaviors, such as gesture, initiating conversation, responding to others' conversation, and turn-taking in conversation, were manually coded by domain experts for system evaluation using a video coding method (Matthew Schmidt et al. 2012). However, manually coding users' behaviors, especially verbal communication, needs significant time and efforts. In addition, the CVE-based intervention systems with this time-consuming measurement method could not provide real-time feedback to the users. These limitations in measuring users' behaviors in CVEs may be

addressed using an intelligent agent with the capability to automatically gauge the users' performance (e.g., communication, collaboration, etc.) within the system itself.

Intelligent agent technology has been explored to measure task performance and conversation behaviors of TD individuals in collaborative learning environments (Kumar et al. 2007; Nabeth et al. 2005; Walker et al. 2014; Scheuer et al. 2010). Researchers in the collaborative learning area have developed intelligent agents to, first, measure important aspects, such as topic change (Van Rosmalen et al. 2005), learner understanding (Linton et al. 2003), quality of arguments (Scheuer et al. 2010), and learner motivation (Desmarais and Baker 2012), of the collaborative learning interactions, and then, provide feedbacks to the users based on the measurements. Although these systems were not designed for ASD intervention, they provided useful information about applying intelligent agent technology to measure user behaviors in CVEs. In this chapter, we present the design of an intelligent agent that could communicate and play games with children with ASD in a CVE in order to measure their communication and collaboration skills.

The main challenge of designing such an intelligent agent is to understand human language using a computer program. It is to be noted 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(Kopp et al. 2005; Cauell et al. 2000). Existing intelligent agents with conversation capabilities could only work in narrowly defined domains (Kopp et al. 2005; Pellom et al. 2001) (Aust et al. 1995). In this work, we also set our goal to design an intelligent agent to communicate and play games with a child with ASD in a narrowly defined domain.

4.2.1. Related work

4.2.1.1. Intelligent agents with conversation capabilities

Intelligent agents with conversation capabilities have been studied for a period of time. One of the early systems in this area, ELIZA, was designed by Weizenbaum in 1966 (Weizenbaum 1966). 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. During the last decades, 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 that can engage conversations using 40000 predefined rules (Shawar and Atwell 2005). 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 (Aron 2011), are typically designed to answer general questions based on predefined question-answer pairs or on-line searching. They could not be directly used for a specific domain, such as game playing, because of 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 (Pellom et al. 2001), train information tracking (Aust et al. 1995), and for museum guide (Kopp et al. 2005). However, there is no common way to design these systems in narrowly defined domains (Allen et al. 2001). These systems varied in their developmental methods considering different purposes, methods to understand linguistic meaning, complexity, robustness, and coverage of domains (Glass 1999; McTear 2002; Eskenazi 2009). 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 the intelligent agents with conversation capabilities for game playing.

4.2.1.2. Intelligent agents with conversation capabilities for game playing

Intelligent agents with conversation capabilities for game playing usually were designed to assist humans in interactive-games. One of the important applications in this area is Non-Player Character (NPC) with conversational capability. The adventure game, Zork-series (Brusk and Lager 2007), 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 (Magerko et al. 2004). Although NPCs in these systems can support communication with players, the communication usually is less-flexible with

fixed-format. Generally such fixed-format methods are not suitable for measuring flexible communication between users in collaborative games.

Only a few intelligent agents with conversation capabilities have been designed to support and measure flexible conversations within the collaborative game domain. Cuayahuitl and colleagues designed an artificial intelligent agent that can play a strategic board game, called Settlers of Catan (Cuayáhuitl et al. 2015). In the board game, players can offer resources for 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 (Kulms et al. 2015). In the collaborative puzzle game, the agent can work together with a human to place blocks of various shapes in three steps: i) one player, either the agent or the human, recommends one of two blocks to the other player, ii) the other player either accepts the recommendation and places the recommended block, or rejects the recommendation and chooses a different block, and iii) 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, very little results have been reported to date about the agent. These technologies provide important guidance about how to design intelligent agent to conduct conversations with a human and measure their communication behaviors. However, they were designed for TD population, and could not be directly used for ASD intervention.

In what follows, we describe the development of our intelligent agent that could communicate with children with ASD and play collaborative puzzle games with them, as well as generate meaningful features to measure their communication and collaboration skills in a CVE. In Section 4.3, we briefly describe the CVE, where the intelligent agent interacted with the children. In Section 4.4, the intelligent agent is described in detail with emphasis on its dialogue manager component. Section 4.5 provides

information about a preliminary study to evaluate the intelligent agent. Results and discussions are presented in Section 4.6. Section 4.7 discusses the limitations of the current work and its possible future extensions.

4.3. Collaborative Virtual Environment

A Collaborative Virtual Environment (CVE), named CoMove, was developed in Chapter II in order to understand and enhance communication and collaboration of children with ASD. The CVE enabled two users to communicate and play collaborative puzzle games in a shared environment. Collaborative puzzle games were selected as the interactive activities in the CVE since these games have been widely accepted for encouraging communication and fostering collaboration in children with ASD (Battocchi et al. 2010). In CoMove, we developed seven collaborative puzzle games in order to stimulate abundant communication and collaborative interactions between users within the system. Fig. 12 shows one example of the games.

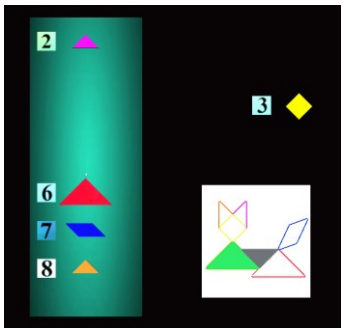

Fig. 12 A collaborative puzzle game in the CVE

Table 14 Key Features and Their Values in Each Collaborative Puzzle Game

Game name	Color visibility	Piece translation control
T1	Both users	One by one
T2	Both users	One by one
T3	Bother users	Together
T4	User1	User2
T5	User2	User1
T6	Bother users	Together
T7	Bother users	Together

The variation of the collaborative puzzle games in the CVE was implemented using two game features, i.e. color visibility and piece translation control. The color visibility feature could encourage users to

share color information; while the piece translation control feature could enable both turn-taking and simultaneous interactions. These collaborative interactions, i.e., information-sharing, turn-taking, and simultaneous interactions, are important for the targeted population and may be related to real time social skills of children with ASD (White et al. 2007). In order to complete these collaborative puzzle games, users were required to communicate with each other to exchange game information and synchronize their game actions. Table 14 shows all the values of these features in each collaborative puzzle game. Take T1 game for example. Two users can see all the colors of puzzle pieces, and they need to move these puzzle pieces one by one in T1 game. The detailed information about the CVE and these collaborative puzzle games can be found in Chapter II.

4.4. Intelligent Agent

4.4.1. Overall description and architecture

We designed an **I**ntelligent agent with the capability of **CO**mmunicatio**N** and **CO**llaboratio**N** (ICON2) in order to measure communication and collaboration skills of children with ASD in the CVE while they played collaborative games. The overall functional view of ICON2 is shown in Fig. 13. ICON2 could perceive a human's speech and game-related actions, i.e. what the human-partner said and what he/she did in the CVE. Then, it generated speech and game-related actions based on the perceived information in a controller. Finally, it executed these generated speech and game-related actions as responses to the human.

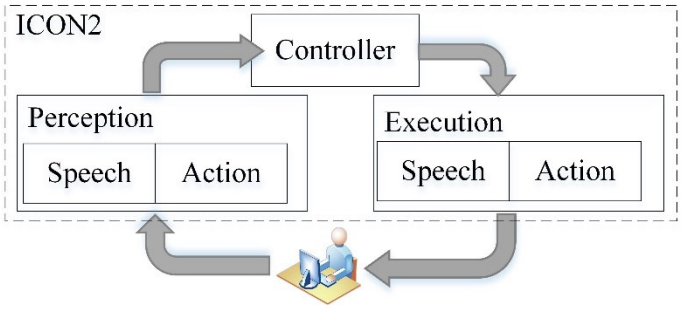

Fig. 13 Overall view of ICON2

The architecture of ICON2 is shown in Fig. 14. 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. The ASR module perceived human's speech inputs; while the TTS module executed speech responses. The GO module perceived game-related information from the CVE; while the AA module executed game-related actions. The DM module was the main component of ICON2, which generated speech and game-related responses based on perceived speech and game-related inputs. The interpretation model and speech lexicon databases were used to help the DM module generate appropriate responses.

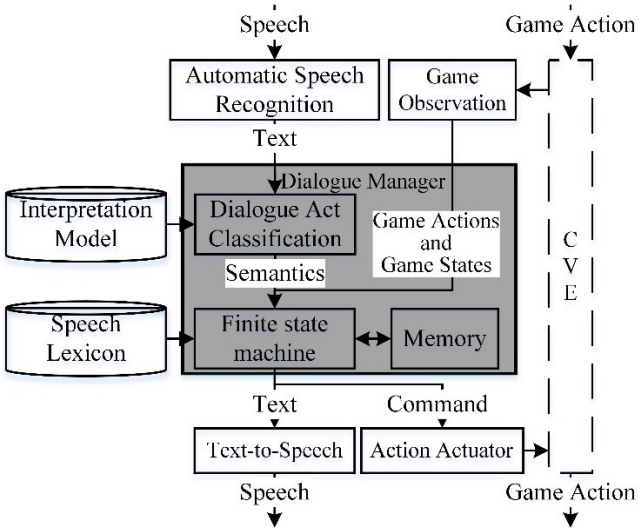

Fig. 14 Architecture of ICON2

Each module of ICON2 was designed in order to support domain-related conversation and collaborative interactions in the CVE. The purpose of the ASR module was to transcribe human speech into text. Google Cloud Speech API[4] was utilized in the ASR module because its low word error rate, i.e., 8%. The GO module could extract game related information, including human's game actions and current game states, from the CVE. The DM module was implemented using a hybrid method, which combined a dialogue act classifier and a finite state machine. The dialogue act classifier classified a human language

[4] cloud.google.com/speech/

into one of several pre-defined categories using an interpretation model shown in Fig. 14. The finite state machine combined speech inputs, game-related inputs, and dialogue history to generate speech and game-related responses. All the dialogue history 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[5] text recognition to transfer the text-based speech presentations to voice responses. The generated game actions could be executed via the AA module. The DM module was the core component of ICON2.

4.4.2. Dialogue manager

Communication and game-playing behaviors of real-users in the CVE in our previous study were analyzed and used for designing the communication and game-playing behaviors of ICON2. In Chapter II, a total of 14 pairs of children, 7 ASD/TD pairs and 7 TD/TD pairs, were involved in playing collaborative puzzle games in a human-human interaction mode in the CVE (Zhang et al. 2016). All the domain-related behaviors of theses real users in the CVE could be presented as pairs of intentions and objects. An intention means an action that a user plans to take when playing a collaborative game. Possible intentions in the CVE include, i) to know the color of a puzzle piece, ii) to provide information, iii) to direct another user to drag a puzzle piece, iv) to acknowledge other's actions, and v) to find a puzzle piece to move. An object means a specific puzzle piece targeted by the intention. Possible values of the object can be any of the seven puzzle pieces or empty. In order to communicate and play games with a real user, ICON2 must be able to i) detect a human's intention and targeted object, and ii) generate appropriate speech and game-related responses based on the detected intention and object.

Besides the communication and game-playing capabilities, ICON2 should also be able to measure users' communication and collaboration skills in the CVE. Therefore, the development of its core component, i.e., the DM module, was required to conform to the following requirements. 1) Since ICON2 was

[5] vuforia.com

required to measure a user's communication and collaboration skills, the DM module must be able to gather or generate features for the measurements. 2) SinceICON2 was required not only to communicate but also to play games in the CVE, the DM module must have the capability to combine both speech and game-related inputs, and generate both speech and game-related responses. 3) As a partner to play collaborative games, ICON2 must be able to act proactively in conversation, i.e., to take initiative, rather than being purely responsive. This means that the DM module must not only respond to user speech but also initiate a conversation.

In order to fulfill these requirements, we developed the communication and game-playing behaviors of ICON2 with three steps: i) understanding a human's spoken natural language and collecting game-related inputs, ii) detecting the human's intention and targeted object from the speech and game-related inputs, and iii) generating speech and game-related responses. 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 measure the users' skills. The second step aimed at combining speech and game-related inputs to enable both communication and game-playing. The third step was important for ICON2 to both respond to the human and initiate a conversation. In summary, the implementation of the DM module took account of all these requirements.

4.4.2.1. Language understanding

We selected combinations of dialogue acts and slots to represent a human's language because such representations were found to be meaningful in measuring communication and collaboration skills of children with ASD. In order to be understandable for a computer, human language is typically represented using a set of messages: each set has a finite number of messages and each message is associated with a particular action (Juang and Furui 2000). One way to represent human utterances is using a set of combinations of dialogue acts and slots. A dialogue act is the specialized performative function that an utterance plays in a language (Stolcke et al. 2000). A slot is a variable that presents specific domain-related information of human utterances (Williams and Young 2007a). Using a combination of dialogue act and slot to represent an utterance has been proven to be useful in previous works (Wen et al. 2015;

Tsiakoulis et al. 2014; Zhu et al. 2014). For example, AT&T spoken dialogue system may represent a caller's request, *I would like to make a payment*, as *Report(payment)*, where *report* is the dialogue act and *payment* is the slot (Gupta et al. 2006).

We defined dialogue acts and slots in our system based on the recorded conversations of our previous study in Chapter II. First, we defined five classes of dialogue acts that were pertinent for the puzzle game, i.e., request_color, provide, direct_movement, acknowledge, and request_object. The descriptions of these dialogue act classes are shown in Table 15. In addition, we defined seven slots, which were color, id, object, action, policy, subject, and out-of-domain, along with several slot words for each slot. The slot words of the first six slots could describe specific features of the collaborative puzzle games. For example, the color slot words, were *red*, *green*, *yellow*, *blue*, *pink*, *orange*, and *gray* that described the color of all the puzzle pieces in the games. The *out-of-domain* slot were used to describe out-of-domain information. Its slot words, such as *name*, *food*, *school*, *weekend*, and *facebook*, were extracted from the out-of-domain utterances in our previous study in Chapter II. The slot words of an utterance were extracted by comparing each word of the utterance with all the predefined slot words; while the dialogue act class of each utterance was computed using an interpretation model.

Table 15 Dialogue Act Classes and Their 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?

We built an interpretation model using conversational data gathered from our previous study in Chapter II, and utilized the model for dialogue act classification during real-time conversation. 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 with the following steps, as shown in Fig. 15. First, we replaced each recognized slot word with its slot type since all the words belonging to a slot perform similar functionality in forming

utterances. This preprocessing procedure was designed to reduce 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 (Fürnkranz 1998; Samuel et al. 1998). Parts of speech and dependency types are also useful structure features in dialogue act classification (Boyer et al. 2010). The Natural Language Toolkit (Bird 2006) was used for 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 were input to train the SVM-RBF model. A 5-fold cross validation was used to select hyperparameters of the SVM-RBF model. The feature extraction method, the PCA model, and the SVM-RBF model were used to classify dialogue act in real time with the same process as shown in Fig. 15.

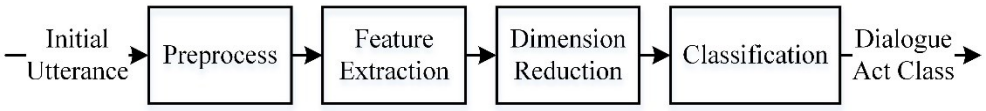

Fig. 15 The process of the online classification

4.4.2.2. Game-related inputs

The game-related inputs, including game actions and current game states, were gathered from the CVE in order for ICON2 to detect the human's intention and object. The partner's game actions were used to represent human behaviors, such as no action for a certain time-duration, dragging a puzzle piece, clicking on a puzzle piece, and releasing a puzzle piece. The current game states were used to represent the interactive environment. The current game states were composed of multiple parameters, and the most important parameters were i) color visibility and ii) piece translation control, which were used to determine the features of each game, as discussed in Section 4.3. Other parameters included color of a puzzle piece, position of each puzzle piece, the target position, and so forth. These game-related inputs were meaningful for ICON2 to detect intention and object.

4.4.2.3. Intention and object detection

We selected a rule-based method to combine the speech and game-related inputs, and to generate speech and game-related responses. In general, spoken dialogue systems that are capable of both speech and non-speech interactions can be implemented using two methods: rule-based and data-driven methods. The rule-based methods updates information and generates responses using predefined rules (Larsson and Traum 2000). Expertise is required to define these rules (DeVault et al. 2011). The data-driven methods, such as reinforcement learning (Williams and Young 2007b), can generate models automatically from training data. However, gathering enough training data is challenging in most cases (Paek and Pieraccini 2008). We have developed a Finite State Machine (FSM) with a set of predefined rules to combine inputs and generate outputs because of the availability of limited training data.

In the FSM, ICON2 combined a partner's speech and game inputs to detect the partner's intention and targeted object, and then, generated speech and game-related responses based on the detected intention and object. When the human-partner spoke to ICON2 or took game actions, the system transferred to the "Intention_Detection" and "Object_Detection" states to detect his/her intention and targeted object. If some information was incomplete, the FSM transferred to the "Intention_Confirm" or "Object_Confirm" states. In these states, ICON2 could seek to clarify unclear information, and gather lost information. After the intention and targeted objects were detected, the system transferred to the "Provide_Information" state to generate responses based on the detected intention and object. In what follows, we present the details of intention detection and object detection in the "Intention_Detection" and "Object_Detection" states.

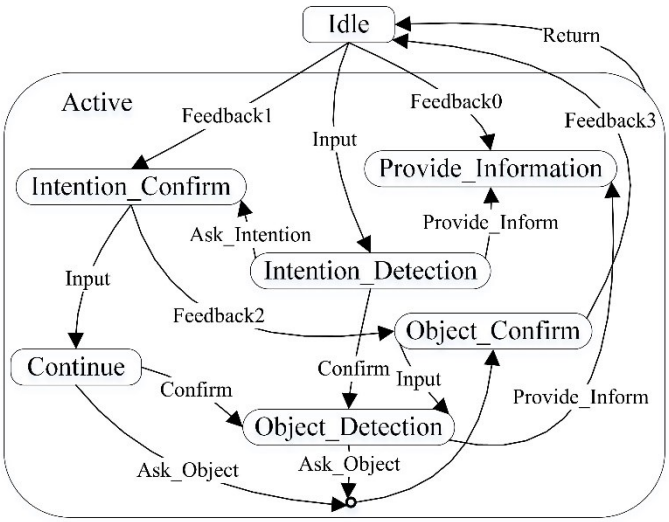

Fig. 16 Finite state machine in the dialogue manager module

The first step of the intention detection was to detect out-of-domain utterances. ICON2 detected out-of-domain utterances based on a rule: if an utterance had *out-of-domain* slot words, the utterance was an out-of-domain utterance. As mentioned in the introduction section, existing spoken dialogue systems were usually designed to operate over a limited and definite domain (Lane et al. 2007). To ensure satisfactory user experience, spoken dialogue system must be able to detect a user's out-of-domain (OOD) utterances, and provide feedback to the user when OOD utterances were detected. Previous literature had applied classification methods to explicitly model OOD utterances for OOD detection (Durston et al. 2001). However, collecting enough training data to model OOD utterances was time-consuming and laborious. Given the limited availability of training data, it was hard to create an OOD model with acceptable accuracy. Therefore, we used a rule-based method to detect OOD utterances in the current study. If the rule-based method fails in detecting an OOD utterance, ICON2 treats the utterance as an in-domain one and asks for additional information to continue playing games. For example, when a user says "*My family went to New York last week*", ICON2 may incorrectly think that the user wants to move a puzzle piece. So, it responds to the user by asking "which puzzle piece do you want to move?" This method turned out to be effective, as discussed in the results session. Other advanced OOD detection methods will be explored in our system in the future.

The intention detection of ICON2 had the advantages of handling ambiguity in natural language. Ambiguity in natural language means that an utterance may have multiple meanings. ICON2 could reduce language ambiguity using game-related inputs and dialogue history based on rules. For example, if a user says "red", she/he may intend *to provide color information* or *to direct ICON2 to move the red puzzle piece*. If the current game state indicates *color being visible for ICON2* or the dialogue history includes *asked for a puzzle piece to move*, the user has a high chance to *direct ICON2 to move the red puzzle piece*. The procedure of intention detection was captured using a tree-structure, as shown in Fig. 17.

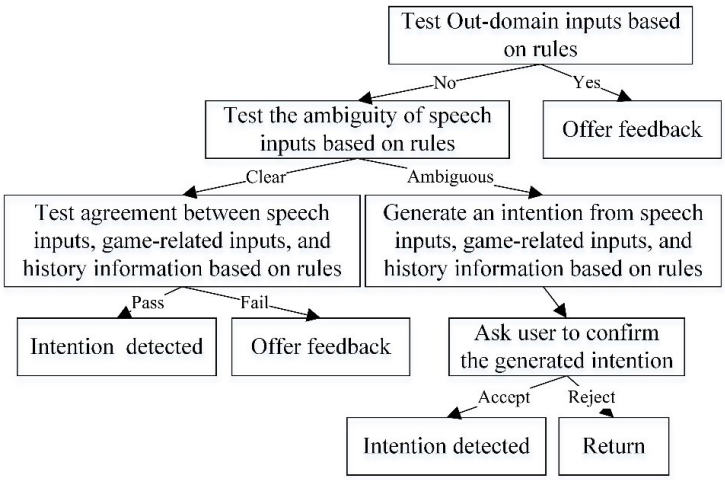

Fig. 17 The logic for intention detection

A weighted average method was developed to combine both speech and game-related information in order to detect the targeted object. Equation (1) computes the similarity between a puzzle piece and the targeted object. A targeted object was usually described using multiple characteristics, such as color of the object, index of the object, and actions on the object. In (1), different characteristics were presented using different terms, such as T_{color}, T_{index}, and T_{active}. The value of each term could be 1 or 0. Each characteristic had a weight, such as W_{color}, W_{index}, and W_{active}, to reflect how important this characteristic was in the object detection. The values of these weights were predefined based on domain knowledge. ICON2 computed a similarity value W_{total} for each puzzle piece based on Equation (1). The object with the highest value was the targeted object. This method has the advantage to handle complex information in conversation using multiple characteristics.

$$W_{total} = W_{color} \times T_{olor} + W_{index} \times T_{index} + \ldots + W_{operation} \times T_{operation} \quad (1)$$

4.4.2.4. Response generation

Based on the detected intention, detected object, and dialogue history, the DM module generated speech and game-related responses using a set of carefully designed IF-THEN rules. One example of the IF-THEN rules is: IF the intention is *out-of-domain*, THEN the agent provides feedback, such as *Hey! I only know something about the game we are playing. Let's play the game!*

ICON2 could not only respond to human conversation but also initiate conversations. The capability to initiate a conversation enabled ICON2 to act proactively in a dialogue, i.e., to take over the initiative, rather than being purely responsive. This led to a more natural conversation. The capability to initiate conversations was implemented using feedback events, such as "Feedback0", "Feedback1", "Feedback2", and "Feedback3", in the FSM (Fig. 16). These events were triggered by game actions and were 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.*"

ICON2 may say different sentences to express the same idea. A speech lexicon was used in order to generate different expressions. The speech lexicon stored multiple expressions for each idea. In real time conversation, ICON2 could 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: i) *what is the color*, ii) *could you tell me the color*, or iii) *is it red or green?* A sample dialogue is shown in Fig. 18.

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 starts moving the puzzle piece).

Silence for a while

Agent: Which puzzle piece do you want to move?

Human: Number six.

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

Fig. 18 A sample dialogue (All game actions are showed in parentheses)

4.5. User Study

A total of five children with ASD, age range: 7 – 17 years, were recruited to participate in a preliminary study to evaluate the communication and collaboration capability of ICON2 in the CVE. All participants had a clinical diagnosis of ASD from a licensed clinical psychologist. All the participants had IQ higher than 70 and were capable of using phrased speech as determined by a trained therapist. The Social Responsiveness Scale, Second Edition (SRS-2) (Constantino and Gruber 2002) and Social Communication Questionnaire Lifetime Total Score (SCQ) (Rutter et al. 2003) were completed by the participants' parents. The characteristics of the participants are shown in Table 16. The experiments were approved by the Vanderbilt University Institutional Review Board (IRB).

Table 16 The Characteristics of the Five Participants

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

During an experiment, a participant played seven collaborative puzzle games with ICON2. After the game playing session, each participant filled out a survey on their opinions about ICON2. We recorded videos of all the experiments. By watching these videos, a human rater rated the participants' improved communication skills and their improved collaboration skills, respectively, using a continuous interval from -4 to 4. The ratings of the improved skills meant how much a participant's communication skills or

collaboration skills improved in the current game compared to the previous game. The ratings were used as the ground truth of the improved skills to evaluate performance of ICON2 in measuring these skills.

4.6. Results and Discussion

Overall, ICON2 worked as designed. All the five participants completed their experiments. Unfortunately, experimental data of a participant in a collaborative puzzle game was lost because the system crashed for unknown reason. We collected data from 34 games ($5 participants \times 7 games - 1 lost = 34 games$). Within the 34 games, a total of 249 utterances were generated by the participants and a total of 374 utterances generated by ICON2.

No out-of-domain utterance has been spoken by the participants. These utterances from the participants were labeled by a human coder as either in-domain or out-of-domain. However, all these utterances were labeled as in-domain utterances, and no one was labeled as out-of-domain utterances. This result was in line with our previous human-human interactions study. In the previous study, a very small percentage, i.e., <0.01, of out-of-domain utterances were spoken by children with ASD when they playing these games with their TD peers.

Results of the dialogue act classification are shown in Table 17. The interpretation model used for the dialogue act classification classified the participants' utterances into five classes: request_color, provide, direct_movement, acknowledge, request_object. A human coder labeled each utterance with one of the five classes, and these labels were used as the ground truth of the classification. The accuracy of the dialogue act classification of ICON2 was 67.47%. This accuracy was higher than the random accuracy of a five-class classifier, i.e., 20%.

Table 17 Dialogue Act Classification Results

		Target class					
		reqcolor	provide	directmove	acknowledge	reqobject	sum
Output class	reqcolor	15	1	2	0	0	18
		6.02%	0.40%	0.80%	0.00%	0.00%	7.23%
	provide	0	93	2	2	0	97
		0.00%	37.35%	0.80%	0.80%	0.00%	38.96%
	directmove	0	14	46	7	0	67
		0.00%	5.62%	18.47%	2.81%	0.00%	26.91%
	acknowledge	0	50	1	14	0	65
		0.00%	20.08%	0.40%	5.62%	0.00%	26.10%
	reqobject	0	0	2	0	0	2
		0.00%	0.00%	0.80%	0.00%	0.00%	0.80%
	sum	15	158	53	23	0	249
		6.02%	63.45%	21.29%	9.24%	0.00%	100.00%

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: i) initiation, which was an utterance used to initiate a conversation, and ii) reply, which was an utterance used to reply to an initiated conversation. 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 213 replies. A human coder labeled each generated utterance as either appropriate or imappropriate. 82.93% of the 161 initiations were labeled as appropriate initiations; while 89.20% of the 213 replies were labeled as appropriate replies. Note that the accuracy of appropriate replies, i.e., 89.20%, was much higher than the accuracy of dialogue act classification, i.e., 67.47%, which suggests that ICON2 could appropriately reply to a human even when it misunderstood the human's language by analyzing the human's game-related inputs, as discussed in sub-section 4.4.2.3.

The results of ICON2 were comparable to other spoken dialogue systems targeted at TD populations. Given the differences in data sample numbers and task domains, it is hard to directly compare numerical results of different spoken dialogue systems. However, we could conclude that the communication capability of ICON2 were comparable to existing spoken dialogue systems by comparing these numerical results. Kopp and colleagues designed a conversational agent as a museum guide to communicate with

museum visitors. The agent could understand a visitor's utterances by mapping keywords using 138 rules. The agent correctly responded to visitors' 50423 utterances with an accuracy of 63% (Kopp et al. 2005). Tewari and colleagues designed a question-answer system (Tewari et al. 2013). The system correctly answered questions with an accuracy of 86%, which were 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 a 84.8% accuracy, which was computed from 46 utterances (Yaghoubzadeh et al. 2015).

The interactions between ICON2 and the participants were comparable to the interactions between two real-users regarding a collaborative movement ratio feature. The collaborative movement ratio is a feature that has been used to measure collaborative efficiency in the CVE (Zhang et al. 2016). It is the time duration ratio of two users simultaneously moving a puzzle piece to an individual user dragging the piece. 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, i.e., 0.11, of children with ASD when they interacted with their TD peers in our previous study in Chapter II.

Results of a distributed survey indicated that children with ASD enjoyed communicating and interacting with ICON2, as shown in Table 18. These participants reported feeling comfortable to talk with ICON2 with an average score of 4 on a 1-5 Likert scale, where 1 means very uncomfortable and 5 means very comfortable. They reported that they could be understood by ICON2 with an average score of 3.8/5 and that they could understand ICON2 with an average score of 4.2/5. It was easy for the participants to play the games with ICON2, as indicated by an average score 4.4/5 on question 5, where 1 means very difficult and 5 means very easy. In addition, they enjoyed playing the games with ICON2 with an average score 4.4/5, where 1 means very dislike and 5 means very like.

Table 18 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 think 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 think you can understand ICON2 very well 1 strongly disagree; 2 disagree; 3 neutral; 4 agree; 5 strongly agree	4.2	0.45
4	Did ICON2 respond to you quickly enough 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 very dislike; 2 dislike; 3 neutral; 4 like; 5 very like	4.4	0.55

Results also indicated that ICON2 had the potential to generate meaningful features to measure communication and collaboration skills of the participants. ICON2 could automatically generate multiple features, shown in Table 19, to represent the participants' behaviors in the CVE. We computed change of a feature, which were the difference of a feature in the current game as compared to the feature in the previous game. Then, we computed the correlation between the change of each feature and the ratings of improved communication skills, as well as the correlation between the change of each feature and the ratings of improved collaboration skills, as discussed in Section 4.5. A spearman's rank correlation indicated a strong correlation ($r_s = 0.72, p < 0.001$) between feature 7 in Table 19 and the ratings of improved communication skills, as well as a strong correlation ($r_s = 0.72, p < 0.001$) between the feature and the rating of improved collaboration skills, as shown in Fig. 19 and Fig. 20.

Table 19 ICON2 Recorded Communication-related Features and Their Descriptions

Index	Feature	Description
1	User word count	How many words the human-partner speaks during a game
2	User utterance frequency	How many times the human-partner speaks during a game
3	Agent word count	How many words the agent speaks during a game
4	Agent utterance frequency	How many times the agent speaks during a game
5	Agent initial frequency	How many times the agent initializes a conversation
6	Agent reply frequency	How many times the agent replies to a conversation
7	Utterance ratio of user and agent	The ratio of number of human's utterance and the number of agent's utterance during a game
8	request_color count	The number of utterance classified as request_color
9	provide count	The number of utterance classified as provide
10	direct_movement count	The number of utterance classified as direct_movement
11	acknowledge count	The number of utterance classified as acknowledge
12	request_object count	The number of utterance classified as request_object

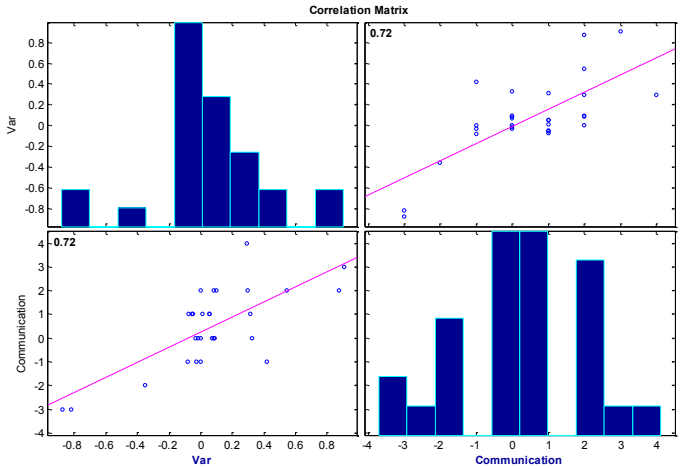

Fig. 19 Correlation between changes of communication/collaboration-related feature 7 and changes of communication skills of children with ASD

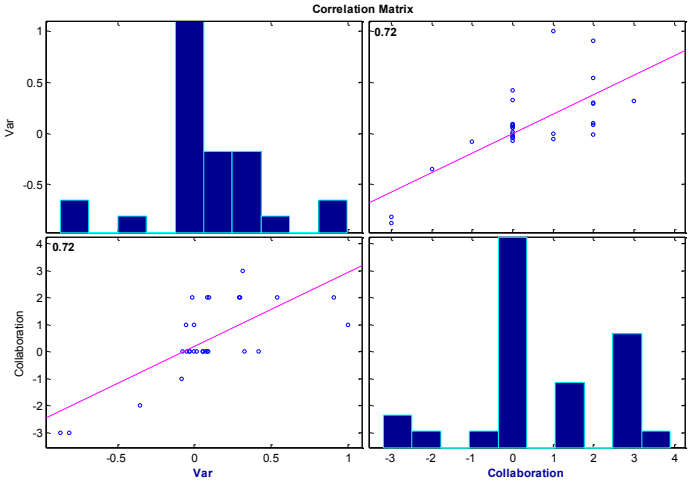

Fig. 20 Correlation between changes of communication/collaboration-related feature 7 and changes of collaboration skills of children with ASD

4.7. Conclusions, Limitations, and Future Work

We designed an intelligent agent, named ICON2, in order to measure communication and collaboration skills of children with ASD by communicating and playing collaborative puzzle games with the children in a CVE. Results of a preliminary study with five children with ASD show that, i) the participants enjoyed communicating and interacting with ICON2 within the CVE; and ii) ICON2 has the potential to communicate and collaborate with children with the participants in the CVE.

ICON2 has the potential to i) appropriately initiate conversation and reply to a participant's conversation; and ii) play collaborative games with the participant. ICON2 generated 82.93% appropriate initiations and 89.20% appropriate replies. These results were comparable to results of other spoken dialogue systems targeted at TD individuals. In addition, the collaborative movement ratio, which was an important feature in collaborative puzzle game as discussed in Chapter II, of the participants when they played the games with ICON2 was comparable to the collaborative movement ratio of participants when they played the games with each other in Chapter II.

Although ICON2 was used in a CVE to play collaborative puzzle games, its communication behaviors could be extended to other domains. In this work, ICON2 extracted meaning of an utterance in the CVE with collaborative puzzle games by mapping the utterance into a dialogue act class (such as request_color, and provide) and a list of slots (such as color, and policy). ICON2 will understand utterances in another domain if the utterances of the domain are used to build a dialogue act classification and define slots. In this study, ICON2 generated speech and operate responses within the CVE based on the detected intention and object. If the intention detection and object detection rules are modified, ICON2 will be able to generate speech and operate responses in another domain.

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 training data used to build the SVM-RBF model for the dialogue act classification was small. The accuracy, 67.47%, of the classifier in this work was much higher than the random accuracy, 20%, of a five-class classifier, and the accuracy, 89.33%, of appropriate responses in this work was comparable to results of other spoken dialogue

systems. These results indicated that the current SVM-RBF model had good performance in dialogue act classification. However, more training data may yield a more accurate classification model.

Second, the keyword-based out-of-domain detection method in this chapter had limitations in detecting out-of-domain utterances. In the current study, the participants did not speak any out-of-domain utterances. Therefore, the limited out-of-domain detection method might be sufficient for the current study. However, future studies should aim to develop more efficient methods for out-of-domain detection.

Third, because of the preliminary nature of this work, the sample size of this study was small. Only five children with ASD were involved to test the intelligent agent, and only 34 data samples were generated for the skills measurements. In future work, we intend to recruit a larger sample of children with ASD and TD children so that we can apply this intelligent agent to measure their skills when they interact with their TD peers.

Despite these limitations, we believe this work contributes to the literature by proposing a novel way to automatically measure both 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/enjoyable to the participants, as well as demonstrated its potential to automatically measure important aspects of interactions in a CVE. This chapter presents the design of the intelligent agent with the capability to communicate and play games with children with ASD. This intelligent agent will be applied to measure both communication and collaboration skills of the children within the CVE in the next chapter.

4.8. References

Allen, J. F., Byron, D. K., Dzikovska, M., Ferguson, G., Galescu, L., & Stent, A. (2001). Toward conversational human-computer interaction. *AI magazine, 22*(4), 27.

Aron, J. (2011). How innovative is Apple's new voice assistant, Siri? *New Scientist, 212*(2836), 24.

Aust, H., Oerder, M., Seide, F., & Steinbiss, V. (1995). The Philips automatic train timetable information system. *Speech communication, 17*(3), 249-262.

Battocchi, A., Ben-Sasson, A., Esposito, G., Gal, E., Pianesi, F., Tomasini, D., et al. (2010). Collaborative puzzle game: a tabletop interface for fostering collaborative skills in children with autism spectrum disorders. *Journal of Assistive Technologies, 4*(1), 4-13.

Benford, S., Greenhalgh, C., Rodden, T., & Pycock, J. (2001). Collaborative virtual environments. *Communications of the ACM, 44*(7), 79-85.

Bird, S. NLTK: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions, 2006* (pp. 69-72): Association for Computational Linguistics

Boyer, K. E., Ha, E. Y., Phillips, R., Wallis, M. D., Vouk, M. A., & Lester, J. C. Dialogue act modeling in a complex task-oriented domain. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 2010* (pp. 297-305): Association for Computational Linguistics

Brusk, J., & Lager, T. Developing Natural Language Enabled Games in (Extended) SCXML. In *Proceedings from the International Symposium on Intelligence Techniques in Computer Games and Simulations (Pre-GAMEON-ASIA and Pre-ASTEC), Shiga, Japan, March, 2007* (pp. 1-3)

Cauell, J., Bickmore, T., Campbell, L., & Vilhjálmsson, H. (2000). Designing embodied conversational agents. *Embodied conversational agents*, 29-63.

Cheng, Y., Chiang, H.-C., Ye, J., & Cheng, L.-h. (2010). Enhancing empathy instruction using a collaborative virtual learning environment for children with autistic spectrum conditions. *Computers & Education, 55*(4), 1449-1458.

Constantino, J. N., & Gruber, C. P. (2002). The social responsiveness scale. *Los Angeles: Western Psychological Services*.

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

Desmarais, M. C., & Baker, R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction, 22*(1-2), 9-38.

DeVault, D., Leuski, A., & Sagae, K. Toward learning and evaluation of dialogue policies with text examples. In *Proceedings of the SIGDIAL 2011 Conference, 2011* (pp. 39-48): Association for Computational Linguistics

Durston, P. J., Farrell, M., Attwater, D., Allen, J., Kuo, H.-K. J., Afify, M., et al. OASIS natural language call steering trial. In *Seventh European Conference on Speech Communication and Technology, 2001*

Eskenazi, M. (2009). An overview of spoken language technology for education. *Speech Communication, 51*(10), 832-844.

Fürnkranz, J. (1998). A study using n-gram features for text categorization. *Austrian Research Institute for Artifical Intelligence, 3*(1998), 1-10.

Glass, J. Challenges for spoken dialogue systems. In *Proceedings of the 1999 IEEE ASRU Workshop, 1999*

Gupta, N., Tur, G., Hakkani-Tur, D., Bangalore, S., Riccardi, G., & Gilbert, M. (2006). The AT&T spoken language understanding system. *IEEE Transactions on Audio, Speech, and Language Processing, 14*(1), 213-222.

Juang, B.-H., & Furui, S. (2000). Automatic recognition and understanding of spoken language-a first step toward natural human-machine communication. *Proceedings of the IEEE, 88*(8), 1142-1165.

Kopp, S., Gesellensetter, L., Krämer, N. C., & Wachsmuth, I. A conversational agent as museum guide–design and evaluation of a real-world application. In *International Workshop on Intelligent Virtual Agents, 2005* (pp. 329-343): Springer

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

Kumar, R., Rosé, C. P., Wang, Y.-C., Joshi, M., & Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. *Frontiers in artificial intelligence and applications, 158*, 383.

Lane, I., Kawahara, T., Matsui, T., & Nakamura, S. (2007). Out-of-domain utterance detection using classification confidences of multiple topics. *IEEE Transactions on Audio, Speech, and Language Processing, 15*(1), 150-161.

Larsson, S., & Traum, D. R. (2000). Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural language engineering, 6*(3&4), 323-340.

Linton, F., Goodman, B., Gaimari, R., Zarrella, J., & Ross, H. Student modeling for an intelligent agent in

a collaborative learning environment. In *International Conference on User Modeling, 2003* (pp. 342-351): Springer

Magerko, B., Laird, J., Assanie, M., Kerfoot, A., & Stokes, D. (2004). AI characters and directors for interactive computer games. *Ann Arbor, 1001*(48), 109-2110.

McTear, M. F. (2002). Spoken dialogue technology: enabling the conversational user interface. *ACM Computing Surveys (CSUR), 34*(1), 90-169.

Millen, L., Hawkins, T., Cobb, S., Zancanaro, M., Glover, T., Weiss, P. L., et al. Collaborative technologies for children with autism. In *Proceedings of the 10th International Conference on Interaction Design and Children, 2011* (pp. 246-249): ACM

Nabeth, T., Razmerita, L., Angehrn, A., & Roda, C. (2005). InCA: a cognitive multi-agents architecture for designing intelligent & adaptive learning systems. *Computer Science and Information Systems, 2*(2), 99-114.

Paek, T., & Pieraccini, R. (2008). Automating spoken dialogue management design using machine learning: An industry perspective. *Speech Communication, 50*(8), 716-729.

Pellom, B., Ward, W., Hansen, J., Cole, R., Hacioglu, K., Zhang, J., et al. University of Colorado dialog systems for travel and navigation. In *Proceedings of the first international conference on Human language technology research, 2001* (pp. 1-6): Association for Computational Linguistics

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*: Western Psychological Services.

Samuel, K., Carberry, S., & Vijay-Shanker, K. Dialogue act tagging with transformation-based learning. In *Proceedings of the 17th international conference on Computational linguistics-Volume 2, 1998* (pp. 1150-1156): Association for Computational Linguistics

Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010). Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning, 5*(1), 43-102.

Schmidt, M., Laffey, J., & Stichter, J. Virtual social competence instruction for individuals with autism spectrum disorders: Beyond the single-user experience. In *Proceedings of CSCL, 2011* (pp. 816-820)

Schmidt, M., Laffey, J. M., Schmidt, C. T., Wang, X., & Stichter, J. (2012). Developing methods for understanding social behavior in a 3D virtual learning environment. *Computers in Human Behavior, 28*(2), 405-413.

Shawar, A., & Atwell, E. (2005). A chatbot system as a tool to animate a corpus. *ICAME Journal: International Computer Archive of Modern and Medieval English Journal, 29*, 5-24.

Stolcke, A., Coccaro, N., Bates, R., Taylor, P., Van Ess-Dykema, C., Ries, K., et al. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics, 26*(3), 339-373.

Tewari, A., Brown, T., & Canny, J. A Question-answering Agent Using Speech Driven Non-linear Machinima. In *International Workshop on Intelligent Virtual Agents, 2013* (pp. 129-138): Springer

Tsiakoulis, P., Breslin, C., Gasic, M., Henderson, M., Kim, D., Szummer, M., et al. Dialogue context sensitive HMM-based speech synthesis. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, 2014* (pp. 2554-2558): IEEE

Van Rosmalen, P., Brouns, F., Tattersall, C., Vogten, H., Bruggen, J., Sloep, P., et al. (2005). Towards an open framework for adaptive, agent-supported e-learning. *International Journal of Continuing Engineering Education and Life Long Learning, 15*(3-6), 261-275.

Walker, E., Rummel, N., & Koedinger, K. R. (2014). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence in Education, 24*(1), 33-61.

Wallace, S., Parsons, S., & Bailey, A. (2015). Self-reported sense of presence and responses to social stimuli by adolescents with ASD in a collaborative virtual reality environment. *Journal of Intellectual and Developmental Disability*.

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.

Wen, T.-H., Gasic, M., Kim, D., Mrksic, N., Su, P.-H., Vandyke, D., et al. (2015). Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking. *arXiv preprint arXiv:1508.01755*.

White, S. W., Keonig, K., & Scahill, L. (2007). Social skills development in children with autism

spectrum disorders: A review of the intervention research. *Journal of autism and developmental disorders, 37*(10), 1858-1868.

Williams, J. D., & Young, S. (2007a). Partially observable Markov decision processes for spoken dialog systems. *Computer Speech & Language, 21*(2), 393-422.

Williams, J. D., & Young, S. (2007b). Scaling POMDPs for spoken dialog management. *IEEE Transactions on Audio, Speech, and Language Processing, 15*(7), 2116-2129.

Yaghoubzadeh, R., Pitsch, K., & Kopp, S. Adaptive grounding and dialogue management for autonomous conversational assistants for elderly users. In *International Conference on Intelligent Virtual Agents, 2015* (pp. 28-38): Springer

Zhang, L., Gabriel-King, M., Armento, Z., Baer, M., Fu, Q., Zhao, H., et al. Design of a Mobile Collaborative Virtual Environment for Autism Intervention. In *International Conference on Universal Access in Human-Computer Interaction, 2016* (pp. 265-275): Springer

Zhu, S., Chen, L., Sun, K., Zheng, D., & Yu, K. Semantic parser enhancement for dialogue domain extension with little data. In *Spoken Language Technology Workshop (SLT), 2014 IEEE, 2014* (pp. 336-341): IEEE

CHAPTER V. APPLICATION OF THE INTELLIGENT AGENT TO MEAUSREMENTS IN A CVE

5.1. Abstract

In Chapter IV, we have designed an intelligent agent in order to measure both communication and collaboration skills of children with ASD in a CVE by communicating and playing collaborative puzzle games with the children within the environment. In this chapter, we present a measurement system that applied the intelligent agent to measure these skills. A preliminary study with 20 pairs of children with ASD and TD children was conducted to evaluate its capability of measuring these skills. Results of the study demonstrated that the system has the potential to generate meaningful features to measure both communication and collaboration skills of the participants when they interacted with the intelligent agent within the CVE. In addition, results of the study indicated that the interactions between the participants and the intelligent agent could reflect important aspects of the interactions between two participants.

5.2. Introduction

In Chapter II, we designed a Collaborative Virtual Environment (CVE), which is a computer-based, distributed, virtual space for multiplayers to interact with one another and/or with virtual items. CVE technology offers a flexible alternative to conventional modalities of both in-vivo (e.g., social skill groups, peer-mediated programs) and technological intervention (e.g., confederate controlled virtual reality (VR), computerized skill programs) where multiple individuals can share and interact in a virtual space using network-based communication. CVEs preserve the advantages of traditional computer-based intervention systems but also facilitate real-time interactions between real users across distance. In particular, the characteristics of this environment are highly controllable and can be adapted and structured in ways that mimic aspects of real-world interactions. These characteristics can tangibly impact the very nature of the collaborative interaction itself.

Although CVEs provide a promising platform for interactions between real users, CVE-based interventions lack reliable and easy-to-use methods for measuring 1) social communication within these systems and 2) impacts of these systems on children with ASD. The majority of CVEs in this area evaluated system impacts based on self-report questionnaires or users' task-performance. For example, Wallace et al. designed a CVE-system to teach greeting behaviors to children with ASD in a virtual gallery (Wallace et al. 2015). They evaluated the system impacts using a self-report questionnaire, and found that children with ASD, compared to their Typically Developing (TD) peers, were less sensitive to a negative greeting. Millen et al. applied CVEs to promote collaboration among children with ASD, and the results of a self-report questionnaire showed improved engagement of children with ASD in the CVEs (Millen et al. 2011). Cheng et al. designed a CVE-based virtual restaurant to understand empathy of children with ASD (Cheng et al. 2010). They found that these children could appropriately answer more empathic questions after the intervention. Although these methods could gather essential information for system evaluation, they could not be used to understand and analyze users' conversation, which is an essential component during user-to-user interactions in most CVE-based interventions.

In some instances, domain experts have been involved to observe and code not only task-performance but also verbal communication of users within CVEs using a human coding methodology. iSocial is a 3D-CVE aimed at understanding and improving social competency development of children with ASD (M Schmidt et al. 2011). In iSocial, children's social behaviors, such as gesture, initiation of conversation, response to others' conversation, and turn-taking in conversation, were manually coded by domain experts for system evaluation using a video coding method (Matthew Schmidt et al. 2012). However, manually coding users' behaviors, especially verbal communication, needs significant time and efforts. In addition, the CVE-based intervention systems, which utilized this time-consuming method for system evaluation, could not provide real-time feedback to the users.

The limitations in measuring social communication in CVEs for pragmatic intervention are due to two fundamental challenges. First, the dynamic social interactions within CVE systems are partner dependent. Quite simply, interactions within the CVE change based on specific partner input and as such

fundamentally limit consistent, controlled, and replicable interactions within the CVE. Second, while open-ended CVE systems pose no restriction in verbal communication between users, subsequent manual coding of interactions is necessary to understand patterns of communication for meaningful measurement and intervention. We believe that these challenges in measuring social communication within CVEs may be addressed using an intelligent agent that can automatically gauge user performance (e.g., communication, collaboration, etc.) within the system itself.

Intelligent agent technology has been explored to measure task performance and conversation behaviors of TD individuals in collaborative learning environments (Kumar et al. 2007; Nabeth et al. 2005; Walker et al. 2014; Scheuer et al. 2010). Note that although designing an intelligent agent that cannot be distinguished from a human for unrestricted naturalistic conversation is a challenge yet to be solved (i.e., the Turing test), designing paradigms for controlling, indexing, and altering aspects of interactions within a specific domain may represent an extremely valuable and much more viable methodology (Kopp et al. 2005; Cauell et al. 2000). Researchers in the collaborative learning area have developed intelligent agents to, first, measure important aspects of the collaborative learning interactions, such as topic change (Van Rosmalen et al. 2005), learner understanding (Linton et al. 2003), quality of arguments (Scheuer et al. 2010), and learner motivation (Desmarais and Baker 2012) and then provide feedbacks to help these users based on the measurements. Although these systems were not designed for ASD intervention, they provided useful information about applying intelligent agent technology to measure the behaviors of the children with ASD in CVEs.

Motivated by this body of work, we designed an intelligent agent that could play collaborative games with children with ASD and provide verbal prompts/responses as it played, within a CVE. At the same time, it generated meaningful features to measure both communication and collaboration skills of the children. Utilizing the intelligent agent as a measurement tool may address existing challenges within this literature.

Collaborative games were selected as interactive tasks in the CVE because they have the potential to facilitate collaboration and communication between users (Leman 2015; Benford et al. 2001). In

particular, collaborative games could create a controllable environment that allow for realistic embodiment of game strategies. As a result, collaborative games with carefully designed strategies could control, and facilitate collaborative interaction between users (Curtis and Lawson 2001; Zancanaro et al. 2007). Battocchi and colleagues designed collaborative puzzle games with an enforced collaboration rule, which required two users to take actions simultaneously to encourage them to work together (Battocchi et al. 2009). They evaluated the effect of these games on users' collaborations by measuring their task-performance, such as task completion time and number of moved puzzle pieces. They found that games equipped with the enforced collaboration rule have more positive effects on children with ASD, compared to these games without these types of rules.

The two primary aims of this work were: i) measuring both communication and collaboration skills of users within a CVE using an intelligent agent; and ii) evaluating whether the measurements could reflect important aspects of peer-mediated interactions in the CVE. In this study, we provided a measurement system that has the potential to measure both communication and collaboration skills of children with ASD in a CVE using an intelligent agent; and conducted a feasibility study with 20 pairs of children with ASD and TD children to evaluate the system. We hypothesized that i) the system has the potential to measure both communication and collaboration skills of children with ASD and their TD peers in human-agent interactions; and ii) the measurements in human-agent interactions could reflect important aspects of the peer-mediated interactions.

In what follows, we present the measurement system to measure both communication and collaboration skills of children with ASD and their TD peers in a CVE using an intelligent agent. Section 5.3 presents the design of the measurement system that applied an intelligent agent to interact with humans in a CVE, as well as a feasibility study used to test the measurement system. Section 5.4 presents a data analysis framework to indicate how the system measured these skills and how we evaluated the system measurements. Results and discussions are presented in Section 5.5. Section 5.6 shows the limitations of the current work and outlines possibilities for future work.

5.3. Method

In this section, we present the design of a system that combined Collaborative viRtual Environment and inTelligent Agent (CRETA) technologies to support both human-agent interactions and human-human interactions. We also conducted a feasibility study with 20 pairs of children with ASD and TD children to test the system. In the next section, we present a framework of data analysis to measure both communication and collaboration skills of these children based on their interactions in CRETA.

5.3.1. System Design

5.3.1.1. Overall description

The measurement system, named CRETA, was aimed at controlling and indexing communication and collaboration behaviors of children with ASD and their TD peers when they play collaborative games in a CVE. CRETA had two components, i.e., a CVE and an intelligent agent. The CVE component was designed for two users to converse and play collaborative games with each other. The puzzle games were equipped with strategies to elicit both communication and collaboration between the two users. The intelligent agent component was designed to interact with humans as well as generate meaningful features to measure their communication and collaboration skills. Fig. 21 shows the architecture of the measurement system. In Fig. 21, *Human_1* and *Human_2* present two humans within the system. These humans used their CVE nodes (*CVE Node_1* and *CVE Node_2*) to interact with each other and with their own intelligent agents, i.e., *Agent_1* and *Agent_2*. The arrows represent data transmission between these components

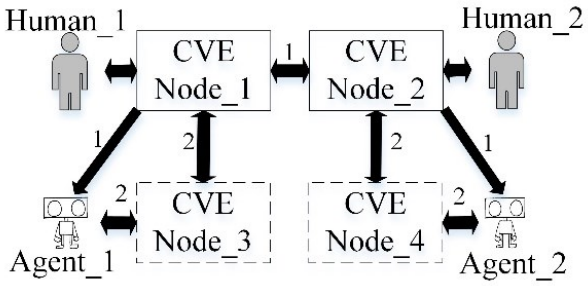

Fig. 21 System architecture

The measurement system provided a platform for both human-agent interactions (HAIs), i.e., interactions between a human and an intelligent agent, and human-human interactions (HHIs), i.e., interactions between two humans. During the HAIs, the agent i) acted as a partner to perform consistent, controlled, and replicable interactions with a child, and ii) measured both communication and collaboration skills of the child through the controlled interactions. During the HHIs, i) the child interacted with his/her peer; and ii) the intelligent agent monitored the interactions between these two peers. The HHI mode was included in order to evaluate whether the measurements in HAIs could reflect important aspects of peer-mediated interactions by comparing a child's behaviors in HAIs to his/her behaviors in HHIs.

5.3.1.2. System components

One important component of CRETA was a CVE, which supported interactions between two users from different locations in a shared environment. Each user, a human or an intelligent agent, utilized a CVE node, which was an instance of the environment, to converse and interact with his/her partner in the shared environment in two ways: conversing via an audio chat functionality, and playing collaborative games. Fig. 22 shows an example of the collaborative games. In addition, the environment could record users' game performance, such as how successfully and how collaboratively one moved the puzzle pieces, as well as task-performance features to represent their collaborative interactions in the CVE. Details about the implementation of the environment are presented in Chapter II.

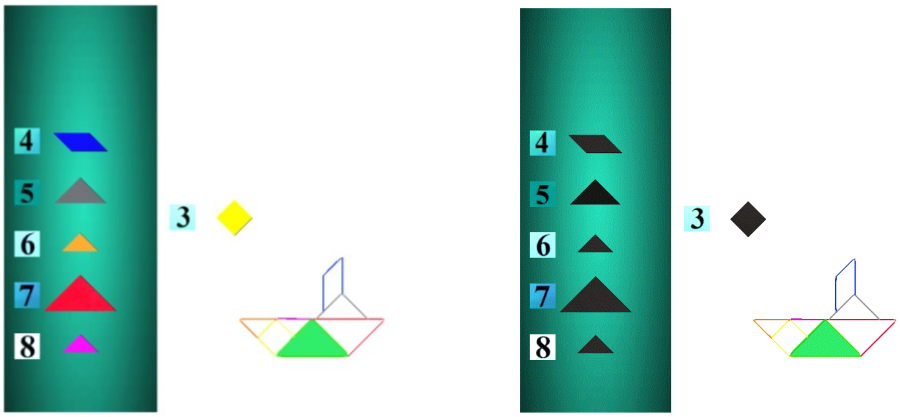

Fig. 22 Environment views of two users (the left image shows the environment view of *Human_1* while the right image shows the environment view of *Human_2*)

We designed multiple collaborative puzzle games with different collaborative strategies in order to elicit communication and collaboration between the users. These collaborative strategies were implemented by manipulating three game features: i) who can see the color; ii) who can move the puzzle pieces; and iii) whether the target area is moving or stationary. The characteristics of these games are shown in Table 20. Take Game_9 for an example: two users need to drag puzzle pieces together to a moving target area, and only one user can see the color of puzzle pieces. Therefore, two users are required to converse with each other to share color information as well as to synchronize their actions in this game.

Table 20 The Features of Each Collaborative Puzzle Game

Game Name	Who can move the puzzle pieces	Who can see the color	Whether the target is moving
Game_1	One by one	Both users	No
Game_2	One by one	One user	No
Game_3	One by one	Another user	No
Game_4	Two users together	Both users	No
Game_5	Two users together	One user	No
Game_6	Two users together	Another user	No
Game_7	Two users together	Both users	Yes
Game_8	Two users together	One user	Yes
Game_9	Two users together	Another user	Yes

Another important component of CRETA was an intelligent agent that could not only converse and play games with a human but also generate meaningful features to reflect his/her communication and collaboration skills. The intelligent agent is a computer program developed using machine learning and natural language processing technologies. When communicating with humans, the intelligent agent, first, transcribed their speech to text in real time using a speech recognition software, i.e., Google Cloud Speech APIs (https://cloud.google.com/speech/). Then, it understood the human language using a dialogue act classification. At the same time, it extracted users' game information, including the human's game actions and current game states, from the CVE. After that, the intelligent agent combined humans' natural language and game information to understand human behaviors. Finally, based on its understanding, the intelligent agent generated speech and game-related responses using a finite state machine. In this procedure, the transcribed text and classified dialogue acts were recorded as verbal-

communication features to present the communication between users. Details of how to design the intelligent agent are presented in Chapter IV.

5.3.1.3. Data transmission between components

These system components communicated with each other by exchanging data. As shown in Fig. 21, CVE node_1 and CVE node_2 transferred audio data (i.e., what Human_1 and Human_2 said), game actions (i.e., what these humans did), and game states (i.e., what kind of game they were playing) between two humans, so that human-users in different locations could converse and play games with each other. In addition, these CVE nodes needed to transfer the audio data and game information to the intelligent agents in order for the intelligent agents to monitor human behaviors in HHIs. The data transmission between CVE node_1 and CVE node_3, and the data transmission between CVE node_2 and CVE_node_4 were used for humans to converse and play games with their intelligent agents within the CVE. All of the data transmission was implemented with socket programming.

By enabling and disabling different kinds of data transmission, CRETA could switch between HHIs and HAIs. In Fig. 21, when the data transmission labeled with number 2 was enabled and the data transmission labeled with number 1 was disabled, each human could communicate with his/her intelligent agent. As a result, a human could converse and play games with his/her intelligent agent. When the data transmission labeled with number 1 was enabled and the one with number 2 was disabled, two humans could communicate with each other. Under this condition, two humans conversed and played games with each other. In the meantime, their audio, game action, and game states data were transferred to the intelligent agents for the intelligent agents to monitor the HHIs.

5.3.2. Feasibility study

We conducted a feasibility study with 20 age- and sex- matched pairs, in order to evaluate whether the intelligent agent could measure both communication and collaboration skills of children with ASD and their TD peers in the CVE.

5.3.2.1. Participants

20 age- and sex- matched pairs were recruited from an existing clinical research registry. Each pair included a child with ASD and a TD child. Participants with ASD had diagnoses from licensed clinical psychologists based upon DSM-5 criteria as well as Autism Diagnostic Observation Schedule-2 scores (Pruette 2013). Additional inclusion criteria included the use of spontaneous phrase speech and IQ scores higher than 70 as recorded in the registry. The IQ criterion was established as a rough proxy for the estimated 5th grade reading level necessary for understanding/completion of the instructions of the CVE tasks. Participants in the TD group were recruited through an electronic recruitment registry accessible to community families. To index initial autism symptoms and screen for autism risk among the TD participants, parents of all participants completed the Social Responsiveness Scale, Second Edition (SRS-2) (Constantino and Gruber 2002) and the Social Communication Questionnaire Lifetime (SCQ) (Rutter et al. 2003). Table 21 shows the characteristics of these participants. The study was approved by the Vanderbilt University Institutional Review Board (IRB).

Table 21 Participant characteristics

	Age	Gender Female/male	SRS-2 total raw score Mean (SD)	SRS-2 T score Mean (SD)	SCQ current total score Mean (SD)
ASD (N=20)	13.33(2.12)	4/16	102.45(23.73)	77.75(9.35)	22.58(8.87)
TD (N=20)	13.50(2.30)	4/16	27.4(21.68)	47.65(8.45)	3(4.08)

Note: SD means standard deviation

5.3.2.2. Experimental Procedure

Each ASD/TD pair completed a one-visit experiment. The procedure for the experiments is shown in Fig. 23. At the very beginning of the experiments, participants were shown an introduction about how to play games in the CVE. Then the participants played nine collaborative puzzle games in a random order. Each game was played in a HHI mode followed by a HAI mode. In the HHI mode, a child with ASD played a game with a TD child for a certain time. Game_1, Game_2, and Game_3 were played for one minute each. Game 4, Game_5, and Game_6 lasted for two minutes each. Each of the other games was played for three minutes. The duration of each game was determined based on our previous study, where four ASD/TD pairs were recruited to play each game without time limitation. In a HAI mode, each child

played the same game with his/her intelligent agent. Between two different games, there was a 10-second break. Each experiment lasted approximately 40 minutes. Participants' behaviors and their computer screens were audio and videotaped during the experiments.

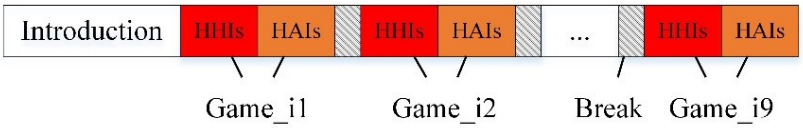

Fig. 23 Experimental procedure

5.4. Data analysis

We present a framework to measure both communication and collaboration skills of the participants in HAIs. The framework measured these skills in three steps, which are shown by the solid lines in Fig. 24. First, the system automatically generated verbal-communication and task-performance features to represent the behaviors of these participants in HAIs. Second, in the *feature evaluation* step, we evaluated whether the system could accurately generate these features, as well as whether these features could reflect important aspects of the behaviors of these participants in HAIs. Third, the system-generated features were then used to measure both communication and collaboration skills in HAIs. In addition, the framework has a *Ground Truth Generation* step, which generated ground truth of the features and the skills.

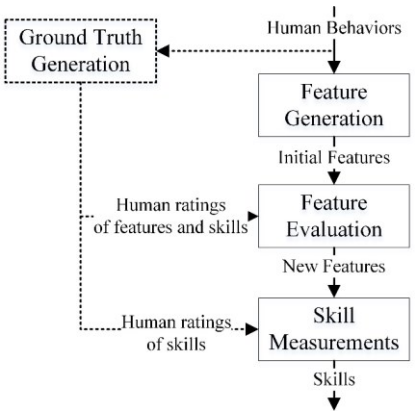

Fig. 24 A framework of data analysis. The solid lines show the procedure to measure communication skills and collaboration skills; while the dotted lines show the procedure to evaluate the measurements

In addition to the framework, we evaluated whether behaviors of the participants in HAIs could reflect important aspects of their behaviors in HHIs. First, we evaluated the system-generated features of HHIs using the method in the *feature evaluation* step of the framework. Then, we computed correlations between each feature of HAIs and the same feature of HHIs. These correlations could demonstrate relationships between their behaviors in HAIs and their behaviors in HHIs.

5.4.1. System-generated features

System-generated features based on previous literature were selected to represent participants' within-system behaviors. All the features and their descriptions are shown in Table 22. The first seven features in Table 22 are verbal-communication features, which were used to represent conversations of the participants. Hourcade and colleagues found that word frequency and sentence frequency could reflect the engagement of children with ASD in collaborative games (Hourcade et al. 2013). Dialogue act features, such as requests for information (McManus and Aiken 1995), providing information (Gogoulou et al. 2008), and acknowledging other people's actions (Vieira et al. 2004), have been proven to be useful in understanding group discussion behaviors of children with ASD and TD children. Task-performance features included how many puzzle pieces have been successfully moved per minute (named *success frequency*), how many times a participant failed to move puzzle pieces (named *failure frequency*), how long he/she dragged puzzle pieces (named *dragging time*), and how often two users collaboratively moved puzzle pieces together (named *collaboration time*). Bauminger-Zviely and colleagues found that the *success frequency* and *failure frequency* features reflected important aspects of collaborative behaviors of children with ASD in collaborative games (Bauminger-Zviely et al. 2013). White and colleagues reported that the *dragging time* and *collaboration time* features could reflect collaboration efficiency of children with ASD when they played collaborative games with their TD peers (White et al. 2007). All these task-performance features were collected by the system in real time; while the verbal-communication features were generated by the intelligent agent using machine learning and natural language processing technologies.

Table 22 system-generated features and their descriptions

Index	Name	Description	Note
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	An example of asking color: *what's the color of this piece*
3	Provide frequency	How many times per minute a user provides game information	An example of providing game information: *this is a red piece*
4	Direct_movement frequency	How many times per minute a user directs movements	An example of directing movements: *move number three*
5	Acknowledge frequency	How many utterances belong to acknowledgements	An example of acknowledgements: *okay*
6	Request_object frequency	How many times per minute a user asks for objects	An example of asking for objects: *which one do you want to move*
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	--

The procedure to generate verbal-communication features included two steps. First, the system automatically detected when spoken sentences ended and transcribed these sentences into text, using the Google Cloud Speech APIs. The *word frequency* and *sentence frequency* features were generated based on the speech recognition results. Then, each sentence was classified into one of the predefined categories, i.e., *Request_color*, *Provide*, *Direct_Movement*, *Acknowledgement*, and *Request_Object*, using a dialogue act classification (see Chapter IV for more information). The corresponding features, such as how many times a participant asked for color information, i.e., *Request_color frequency*, how many times the participant provided game information for his/her partner, i.e., *Provide frequency*, and how many times the participant directed his/her partner to move puzzle pieces, i.e., *Direct_Movement frequency*, were computed based on the classification results.

The verbal-communication features and task-performance features, which were generated in different system components and computers, were synchronized using multiple synchronization methods for

offline data analysis. The intelligent agent component, including Agent_1 and Agent_2 in Fig. 21, generated and recorded verbal-communication features; while the CVE component, including CVE Node_1, Node_2, Node_3, and Node_4 in Fig. 21, generated and recorded task-performance features. These system components were located on two desktop computers. Specifically, the CVE Node_1, Node_3 and Agent_1 were run on one computer, while CVE Node_2, Node_4, and Agent_2 were run on another computer. The features on the same computer were synchronized using timestamps corresponding with these features; while the features on two different computers were synchronized using *game start times*. The *game start times* of two computers were recorded and synchronized at the beginning of each game in real time. Because all the features generated during a game were grouped as one data sample, these synchronization methods were sufficient for our data analyses.

5.4.2. Human Ratings

Two human raters, a primary human rater and a secondary human rater, watched videos of the experiments, and rated both communication and collaboration skills of the participants in order to provide the ground truth of these skills. The primary human rater was blinded to the study; while the secondary human rater was not blinded to the study. The primary human rater rated all the experiments, and her ratings were used as the ground truth of the skills. The secondary human rater rated 25% of the all the experiments, and her ratings were used to evaluate the human rating results. We selected the human ratings of these skills as the ground truth because the goal of this study was to replace the time-consuming human-rating method in measuring these skills.

These two human raters utilized the same rating scheme, and rated the skills independently. The human raters rated these skills along two kinds of rating scales: i) a binary rating, which has a value 1 or 0, and ii) a continuous rating, which has a value between -4 and 4. Values of the binary rating indicated whether the raters felt participants had high levels or low levels of communication skills or collaboration skills in the current game. Values of the continuous rating showed how good the participants' skills were in the current game.

The inter-rater reliability, which indicated the degree of agreement between two human raters in their ratings, was assessed in two ways. First, inter-rater agreement on the binary ratings was assessed using a Cohen's Kappa method, which is suitable to assess inter-rater agreement for categorical items of two coders (Eugenio and Glass 2004). Agreement on a binary rating in this study means both human raters rated skills of a participant as a high level or both of them rated the skills as a low level. Second, the inter-rater reliability of continuous rating was assessed using the Spearman's rank correlation (Mathiowetz et al. 1984) to indicate the relationship of the continuous ratings between two human raters.

A human rater, different from the previous two human raters, manually rated participants' conversations from 20% of experimental sessions in order to provide ground truth for system-generated verbal-communication features. This human rater was a native English speaker and was blinded to the study. Specifically, the rater watched videos recorded during the experiments and manually transcribed the participants' speech to text. Then the rater labeled each sentence with one of the predefined dialogue acts, i.e., *Request_Color*, *Provide*, *Direct_Movement*, *Acknowledge*, and *Request_Object*. The manually transcribed texts and the labels assigned to these sentences were used as the ground truth to evaluate the system-generated verbal-communication features.

5.4.3. Feature processing

We preprocessed system-generated features by removing outliers and normalizing these feature values. Statistical tests and machine learning methods are sensitive to the outliers in data samples. Therefore, we removed outliers using a univariate method, which removed data samples that have extreme values on one feature (Grubbs 1969). Then, each feature of the data samples was normalized using a min-max normalization method (Jain and Bhandare 2011) to allow comparisons across different features. After the preprocessing procedure, we evaluated i) whether the system could accurately generate verbal-communication features; and ii) whether the system-generated features could reflect important aspects of both communication and collaboration skills of the participants.

We analyzed performance of the system in generating verbal-communication features. As discussed in Section 5.4.1, the intelligent agent generated the verbal-communication features using a speech recognition software and a dialogue act classification model. We evaluated the performance of the speech recognition software using its word error rate (Klakow and Peters 2002). Then, we evaluated the performance of the dialogue act classification model by showing its confusion matrix, which included true positives, true negatives, false positives, and false negatives of each dialogue act class (Srinivasan and Petkovic 2000). Finally, we computed error rates of system-generated verbal-communication features. A feature error rate is a ratio of a feature error to its true value. The feature error is the difference between the measured feature and its true value. We also computed the ratio of the sentence number of each dialogue act to the total number of sentences. The ratio helped describe the error rate of the corresponding verbal-communication feature. Results of the analysis indicated whether the system could accurately generate the verbal-communication features.

To evaluate whether the system-generated features could reflect important aspects of both communication and collaboration skills, we computed correlations between the system-generated features and the human ratings of communication skills on a continuous scale, as well as correlations between system-generated features and the human rating of collaboration skills on a continues scale. We selected Spearman's rank correlation, which is a non-parametric measure of rank correlation between two variables (Krishnaiah 1980), to compute the correlation because these features did not follow a normal distribution. It has been commonly accepted that if the correlation between a feature and the skills is between -.3 and .3, the feature has a small strength of association with the skills (Cohen 1988). If the correlation is between .3 and .5 or between -0.5 and -0.3, the feature and the skills have a moderate strength of association. Otherwise, the features and the skills have a strong correlation. We also computed correlations between each feature and the human ratings of the skills on a binary scale using a Rank Biserial correlation, which is used to find a correlation between binary nominal data and ranked data (Cureton 1956).

5.4.4. Skill measurements

We built machine learning models to measure both communication and collaboration skills using the system-generated features. In particular, we trained machine learning models to classify a data sample, which includes all system-generated features of a game, into a binary-class, i.e., a high level of skills or a low level of skills. We selected Support Vector Machine with Radial Basis Function (SVM-RBF) kernel as the machine learning methods for the classification given the fact that SVM-RBF methods usually have good performance for classifying data with a small sample size (Chang et al. 2010). A SVM-RBF model was built to measure communication skills using the system-generated features and ratings of communication skills on a binary scale; while another SVM-RBF model was built to measure collaboration skills using these features and rating of the collaboration skills on a binary scale. In addition, we trained two models to classify these skills based on balanced training data. The balanced training data were generated by randomly under-sampling the majority class, which is a commonly used resampling techniques to improve classification performance in unbalanced datasets. 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.

5.5. Results

5.5.1. Human-Agent interaction results

We present results of each step of the framework to measure both communication and collaboration skills of the participants in HAIs. In Section 5.5.1.1, we present results of inter-rater reliability regarding human ratings of the skills in HAIs to indicate whether the human ratings were reliable. Then, we show feature evaluation results to indicate i) whether the system could accurately generate verbal-communication features, and ii) whether the system-generated features could reflect important aspects of behaviors of the participants in the HAIs. Finally, we provide results of measuring both communication and collaboration skills from the system-generated features in HAIs.

5.5.1.1 Results of human ratings of HAIs

We computed inter-rater reliability regrading human ratings of communication skills and human ratings of collaboration skills in HAIs. The inter-rater agreement on the human ratings of communication skills on a binary scale in HAIs was 85.50%; while the inter-rater agreement on the human ratings of collaboration skills on a binary scale in HAIs was 77.96%. Regarding the human ratings of communication skills on a continuous scale in HAIs, we found a moderate correlation (r_s=0.42, p<.001) between two human raters. Regarding the ratings of collaboration skills on a continuous scale in HAIs, Spearman's rank correlation indicated a strong correlation (r_s=0.54, p<.001) between the two human raters.

5.5.1.2. Results of feature evaluation of HAIs

We tested whether the system could accurately generate verbal-communication features. The word error rate of the speech recognition was 18.01% in HAIs. In HAIs, the accuracy of the five-class dialogue act classification was 70.27%, 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 23. These accuracies were computed based on 1337 spoken sentences of the participants.

Table 23 dialogue act classification accuracies in HAIs

		Classification results					
		Request_ color	Provide	Direct_ movement	Acknowledge	Request _object	Sum
Expected 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%

The error rate[6] of each verbal-communication feature in HAIs are shown in Table 24. The *setence frequency* feature in HAIs had the lowest error rate, 0.0566. This result indicated that the system has the potential to accurately generate the *sentence frequency* feature in HAIs. However, the *Request_color frequency* and *Request_object frequency* feature had high error rates in HAIs. Because of the high error rates, we removed these features for the following data analysis.

Table 24 Error rate of each system-generated feature in HAIs

System-generated Feature	Error rate in HAIs	Ratio of the number of sentences belonging to 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	--

Note: error rate means the ratio of incorrectly detected features to the true value of the features. It may be larger than 1.

We computed correlations between system-generated features and continuous ratings of communication skills, as well as correlations between system-generated features and continuous ratings of collaboration skills, when participants interacted with the intelligent agents. In Table 25, the 2nd column (named *correlation between a feature and continuous communication skills in HAIs*) shows the correlations between each system-generated feature and the ratings of communication skills on a continuous scale in HAIs. The 3rd column (named *correlation between a feature and continuous collaboration skills in HAIs*) shows the correlations between each system-generated feature and ratings of collaboration skills on a continuous scale in HAIs. Regarding the continuous ratings of communication

[6] The error rate may be larger than 1. Take word frequency for example. If a participant says "*one*" and the system detects "*one and*", the system incorrectly detects one word and the error rate is 1. Note that the system correctly detects one word even when the error rate is 1. The error rates of other features depended on errors of both speech recognition and dialogue act classification, which are presented at the very beginning of Section 5.5.1.2.

skills in HAIs, although many correlations were significant, none were considered strong. Regarding the continuous ratings of collaboration skills in HAIs, there was a negative strong correlation between the ratings and *failure frequency* (r_s=-0.5487, p<.001).

Table 25 Correlation between a system-generated feature and human ratings on a continuous scale in HAIs

System-generated feature	Correlation between a feature and continuous communication skills in HAIs	Correlation between a feature and continuous collaboration skills in HAIs
Word frequency	0.3804**	0.1536*
Provide frequency	0.4294**	0.3403
Direct_movement frequency	0.2186**	-0.0069
Acknowledge frequency	0.0292	0.0382**
Sentence frequency	0.3541**	0.2288**
Success frequency	0.4039**	0.4091**
Failure frequency	-0.4479**	-0.5487**
Collaboration time	0.3540**	0.3528**
Dragging time	0.1844*	0.1269*
Collaborative movement ratio	0.3369**	0.3538**

Note: ** indicates a p value less that .001; * indicates a p value less than .05

The correlations between each system-generated feature and the ratings of skills on a binary scale in HAIs are shown in the 2nd column and 3rd column of Table 26. Regarding the binary ratings of communication skills in HAIs, moderate correlations were found for *Provide frequency* (r_{rb}=.3464, p<.001), *sentence frequency* (r_{rb}=.4175, p<.001), *success frequency* (r_{rb}=.3117, p<.001), and *failure frequency* (r_{rb}=-0.3970, p<.001). Regarding the binary ratings of collaboration skills in HAIs, moderate correlations were found for *success frequency* (r_{rb}=.3101, p<.001), and *failure frequency* (r_{rb}=-0.4416, p<.001).

Table 26 Correlation between a system-generated feature and human ratings in a binary scale in HAIs

System-generated feature	Correlation between a feature and binary communication skills in HAIs	Correlation between a feature and binary collaboration skills in HAIs
Word frequency	0.1865**	0.0344
Provide frequency	0.3464**	0.2443**
Direct_movement frequency	0.0628	-0.0804
Acknowledge frequency	0.1221	0.0729
Sentence frequency	0.4175**	0.1594*
Success frequency	0.3117**	0.3101**
Failure frequency	-0.3970**	-0.4416**
Collaboration time	0.2380**	0.2769**
Dragging time	-0.2756*	0.0718
Collaborative movement ratio	0.0456**	0.2782**

Note: ** indicates a p value less that .001; * indicates a p value less than .05

5.5.1.3. Results of skill measurements of HAIs

The system-generated features with SVM-RBF models could assess both communication and collaboration skills in HAIs with high accuracies. The accuracy to assess binary communication skills in HAIs is 76.67% with balanced data samples, i.e., 32 data samples belonging to high levels of communication skills and 32 in low levels. The collaboration skills could be assessed with a high accuracy, 82.14%, with balanced data samples, i.e., 42 data samples belonging to high levels of collaboration skills and 42 data samples belonging to low levels. The accuracies of measuring both communication and collaboration skills with all the data are shown in the Table 27.

Table 27 Accuracies of measuring both communication and collaboration skills

Index	Which skills to measure?	Data sample size (high level / low level)	Accuracy of balanced data	Accuracy of all data
1	Communication skills in HAIs	244/32	82.14%	93.75%
2	Collaboration skills in HIAs	234/42	76.67%	88.69%

5.5.2. Human-Human interaction results

In order to evaluate whether participant behaviors in HAIs could reflect important aspects of their behaviors in HHIs, we i) evaluated system-generated features in HHIs, and ii) compared the features of HAIs to the features of HHIs in this section. In Section 5.5.2.1, we present results of inter-rater reliability of human ratings in HHIs to indicate whether the human ratings in HHIs were reliable. Section 5.5.2.2 shows feature evaluation results to indicate whether the system could accurately generate verbal-

communication features in HHIs, as well as whether the system-generated features could reflect important aspects of participant behaviors in HHIs. Finally, we present correlations between features of HHIs and features of HAIs. These system-generated features were used to reflect important aspects of participant behaviors within the system. The correlations were computed to indicate relationships between participant behaviors in HAIs and their behaviors in HHIs.

5.5.2.1. Results of human rating of HHIs

We computed inter-rater reliability regrading human ratings of communication skills and human ratings of collaboration skills in HHIs. The inter-rater agreement on the human ratings of communication skills on a binary scale in HHIs was 74.47%; while the inter-rater agreement on the human ratings of collaboration skills on a binary scale in HHIs was 87.50%. Regarding the ratings of communication skills on a continuous scale in HHIs, we found a strong correlation (r_s=0.73, p<.001) between two human raters. Regarding the ratings of collaboration skills on a continuous scale in HHIs, Spearman's rank correlation also indicated a strong correlation (r_s=0.78, p<.001) between two human raters.

5.5.2.2. Results of feature evaluation of HHIs

We tested whether the system could accurately generate verbal-communication features in HHIs. The error rate of the speech recognition was 23.16% in HHIs. The dialogue act classification, a five-class classification, could classify participants' spoken sentences with a 68.78% accuracy in HHIs. The detailed results of the dialogue act classification in HHIs are shown in Table 28. These accuracies were computed based on 868 sentences.

Table 28 dialogue act classification accuracies in HHIs

		Classification results					
		Request_color	Provide	Direct_movement	Acknowledge	Request_object	Sum
Expected	Request_color	**1.38%**	0.35%	0.58%	0	0.12%	2.42%
	Provide	0.23%	**33.06%**	15.09	5.53%	0.12%	54.03%
	Direct_movement	0	2.07%	**21.66%**	0.46%	0	24.19%
	Acknowledge	0	3.92%	3.92%	**12.33%**	0	18.66%
	Request_object	0	0.12%	0.23%	0	**0.35%**	0.69%
	Sum	1.61%	38.02%	41.47%	18.32%	0.58%	100%

The error rate of each verbal-communication feature in HHIs are shown in Table 29. The *Word frequency* feature in HHIs had the lowest error rate, 0.1777. This result indicated that the system has the potential to accurately generate the *word frequency* feature in HHIs. However, the *Request_color frequency* and *Request_object frequency* feature had high error rates in HHIs. Because of the high error rates, we removed the features for the following data analysis.

Table 29 Error rate of each system-generated feature in HHIs

System-generated Feature	Error rate in HHIs	Ratio of a frequency to sentence frequency in HHIs
Word frequency	0.1777	--
Request_color frequency	1.4615	0.0110
Provide frequency	0.5339	0.6988
Direct_movement frequency	0.8589	0.1379
Acknowledge frequency	0.4471	0.1438
Request_object frequency	1.2000	0.0085
Sentence frequency	0.1991	1

Note: error rate means the ratio of incorrectly detected features to the true value of the features. It may be larger than 1.

As seen in Table 30, we found strong correlations between several system-generated features and continuous ratings of communication skills in HHIs, as well as strong correlations between several system-generated features and continuous ratings of collaboration skills in HHIs. Regarding continuous ratings of communication skills in HHIs, Spearman's rank correlation indicated strong positive correlations between the ratings and *word frequency* (r_s=.7578, p<.001), *provide frequency* (r_s=.5422, p<.001), *Direct_movement frequency* (r_s=.6673, p<.001), and *sentence frequency* (r_s=.7649, p<.001).

Regarding continuous ratings of collaboration skills in HHIs, although many correlations were statistically significant, the only one considered strong was *success frequency* (r_s=0.5378, p<.001).

Table 30 Correlation between a system-generated feature and human ratings on a continuous scale in HHIs

System-generated feature	Correlation between a feature and continuous communication skills in HHIs	Correlation between a feature and continuous collaboration skills in HHIs
Word frequency	0.7578**	0.2994**
Provide frequency	0.5422**	0.1855**
Direct_movement frequency	0.6673**	0.2908**
Acknowledge frequency	0.3472**	0.1489
Sentence frequency	0.7649**	0.3932**
Success frequency	-0.0843	0.5378**
Failure frequency	-0.0345	-0.3714**
Collaboration time	-0.0294	0.3839**
Dragging time	0.2696**	0.2183**
Collaborative movement ratio	-0.1548	0.3864**

Note: ** indicates a p value less that .001; * indicates a p value less than .05

As seen in Table 31, we used a Rank Biserial correlation to examine relations between each system-generated feature and binary ratings of communication skills in HHIs. Regarding the binary ratings of communication skills in HHIs, strong positive correlations were found for *word frequency* (r_{rb}=.5750, p<.001), *Direct_movement frequency* (r_{rb}=.5249, p<.001), and *sentence frequency* (r_{rb}=.6446, p<.001). Regarding binary ratings of collaboration skills in HHIs, although many correlations were significant, none were considered strong.

Table 31 Correlation between a system-generated feature and human ratings in a binary scale in HHIs

System-generated feature	Correlation between a feature and binary communication skills in HHIs	Correlation between a feature and binary collaboration skills in HHIs
Word frequency	0.5750**	0.2316**
Provide frequency	0.3965**	0.1618*
Direct_movement frequency	0.5249**	0.2540**
Acknowledge frequency	0.2988**	0.1384
Sentence frequency	0.6446**	0.3107**
Success frequency	-0.1462	0.3214**
Failure frequency	-0.0061	-0.3147**
Collaboration time	-0.0060	0.3404**
Dragging time	0.1945**	0.2480**
Collaborative movement ratio	-0.1827	0.2178**

Note: ** indicates a *p* value less that .001; * indicates a *p* value less than .05

5.5.2.3. Correlations between features of HHIs and features of HAIs

A Spearman's rank correlation analysis was used to determine the relationship between a feature of HAIs and the feature of HHIs, as shown in Table 32. There was a strong correlation (r_s=0.6080, p<.001) between the HHIs' *word frequency* and the HAIs' *word frequency*. In addition, there was a strong correlation (r_s=0.5922, p<.001) between the s*entence frequency* feature of HHIs and the *sentence frequency* feature of HAIs. Spearman's rank correlation also indicated a strong correlation (r_s=0.5913, p<.001) between the *collaboration time* of HHIs and the *collaboration time* of HAIs.

Table 32 Correlations between features in HHIs and them in HAIs

System-generated feature1	System-generated feature2	Correlation between system-generated feature1 and system-generated feature 2
Word frequency in HHIs	Word frequency in HAIs	0.6080**
Request_color frequency in HHIs	Request_color frequency in HAIs	0.2749**
Provide frequency in HHIs	Provide frequency in HAIs	0.4463**
Direct_movement frequency in HHIs	Direct_movement frequency in HAIs	0.3366**
Acknowledge frequency in HHIs	Acknowledge frequency in HAIs	0.2765**
Sentence frequency in HHIs	Sentence frequency in HAIs	0.5922**
Success frequency in HHIs	Success frequency in HAIs	0.0562
Failure frequency in HHIs	Failure frequency in HAIs	0.1382
Collaboration time in HHIs	Collaboration time in HAIs	0.5913**
Dragging time in HHIs	Dragging time in HAIs	0.3631**
Collaborative movement ratio in HHIs	Collaborative movement ratio in HAIs	0.2217**

Note: ** indicates a *p* value less that .001; * indicates a *p* value less than .05

5.6. Conclusion, limitations, and future work

We applied an intelligent agent in order to measure both communication and collaboration skills of children with ASD and their TD peers in a CVE. Given the challenges in understanding unrestricted conversation between real-users in the CVE, we designed an intelligent agent that could interact with these children to control their behaviors in the CVE, as well as to automatically measure both communication and collaboration skills of these children through the controlled interactions. Our results indicated that i) the system could measure these skills of children with ASD and their TD peers when they played games with the intelligent agent; and ii) their interactions with the intelligent agent could reflect important aspects of peer-mediated interactions in the CVE.

We found that our system could generate meaningful features to automatically measure both communication and collaboration skills of the participants in HAIs. First, the system could accurately generate verbal-communication features as indicated by the low error rates of these features. For example, the *sentence frequency* feature in HAIs had a low error rate 0.0566. Second, we found a moderate correlation between the word frequency features and human ratings of communication skills in HAIs, as well as a negative strong correlation between the ratings and *failure frequency* (r_s=-0.5487, p<.001) and the continuous ratings of collaboration skills in HAIs. Third, all the features together could measure these skills with high accuracies using machine learning models. The accuracy to measure the communication skills was 82.14%, while the accuracy to measure the collaboration skills was 76.67%. Although these machine learning models were built offline, they could be used for real-time measurements in the future. Therefore, the system has the potential to automatically measure both communication and collaboration skills in human-agent interactions based on these system-generated features.

Some system-generated features in HHIs may reflect important aspects of the peer-mediated interactions. Previous literature found that *word frequency* and *sentence frequency* features could be used to evaluate social communication of children with ASD in collaborative games (Hourcade et al. 2013). Our results were in line with these findings. We found a strong correlation between the *word frequency* feature and the communication skills on a continuous scale in HHIs, as well as a strong correlation between the *sentence frequency* feature and the skills. These results indicated a strong association between each of the features and the skills. Therefore, these features could reflect important aspects of the communication skills of the participants in peer-mediated interactions. We also found a strong correlation between a *success frequency* feature and collaboration skills on a continuous scale in HHIs, as well as a moderate correlation between a *collaboration time* feature and the skills in HHIs. These results may indicate that these features could reflect important aspects of the collaboration skills in peer-mediated interactions. This finding is in line with previous literature, which utilized *success frequency* (Bauminger-Zviely et al. 2013) and *collaboration time* (Wilson and Russell 2007) to evaluate system impacts on the collaboration skills of children with ASD when they played collaborative puzzle games.

The interactions between a human and the intelligent agent may reflect important aspects of peer-mediated interactions. Spearman's rank correlation demonstrated strong positive correlations between each of the system-generated features, i.e., *word frequency*, *sentence frequency*, and *collaboration time*, in HAIs and the features in HHIs. These results support our hypothesis that human-agent interactions could reflect important aspects of the interactions between real-participants.

The errors 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 23 and Table 28. 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 a very few *Request_object* sentences, as indicated by the small ratio of the *Request_object* frequency to the *sentence* frequency in Table 24. As a result, a very few incorrectly detected *Request_object* sentences could lead to a high error rate. We found the same results regarding the *Request_color frequency* feature.

Our system represents a novel contribution to the literature by providing a way to automatically measure both communication and collaboration skills of children with ASD and their TD peers in a CVE. Most existing CVE intervention systems can only automatically generate task-performance features to measure peer-mediated interactions. Verbal-communication behaviors are informative in representing collaborative interactions between peers (Owen-DeSchryver et al. 2008). Previous work has examined verbal-communication behaviors in order to understand peer-mediated interactions (Matthew Schmidt et al. 2012). However, past studies utilized a time-consuming human-coding method for the analysis. Our system could automatically generate meaningful verbal-communication and task-performance features to measure both communication and collaboration skills of the participants within the CVE. Therefore, this

system has the potential to save time, effort, and costs associated with the human coding methodology in measuring these skills of children with ASD and their TD peers in the CVE.

Although these results are promising, there are some limitations in current study. First, the sample size was relatively small, and the experimental design consisted of only one session. Note that the goal of the present study was to design a measurement system to automatically measure important aspects of interactions in a CVE with a preliminary study. Our preliminary results indicated that this system has the potential to automatically measure both communication and collaboration skills of children with ASD and their TD peers in the CVE. In the next step, we will utilize this system for real-time measurements with more participants and a longer intervention duration.

Second, in order to be used as a measurement tool, the measurement system needs to be tested across a range of treatment approaches. This study only tested whether it could measure both communication and collaboration skills in a CVE with collaborative puzzle games. The system in this study was limited in understanding verbal-communication and task-performance within this system. This system will need to be extended to measure interactions in other domains by modifying the dialogue act classification of the intelligent agent. To modify the classification means to train a different classification using conversational data in that domain (see Chapter IV for more information about the classification).

Third, the system-generated features were limited. We only explored 12 features for the measurements in the current study. Human behaviors, such as their eye gaze, body language, and facial expression, could also provide essential 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 with separate algorithms, such as eye gaze recognition, gesture recognition, and emotion recognition, in order to understand the non-verbal communications.

Despite these limitations, this work contributes to the literature by proposing a novel way to automatically measure both communication and collaboration skills of children with ASD and their TD peers within a CVE using an intelligent agent. To the best of our knowledge, this is the first system that could automatically measure these skills. Such a system can reduce time and costs associated with the

traditional human-coding methodologies, as well as enable real-time feedback within an intervention system. As a result, this work, at least partially, addresses the limitations in measuring important aspects of interactions within a CVE.

5.7. References

Battocchi, A., Pianesi, F., Tomasini, D., Zancanaro, M., Esposito, G., Venuti, P., 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* (pp. 197-204): ACM

Bauminger-Zviely, N., Eden, S., Zancanaro, M., Weiss, P. L., & Gal, E. (2013). Increasing social engagement in children with high-functioning autism spectrum disorder using collaborative technologies in the school environment. *Autism, 17*(3), 317-339.

Benford, S., Greenhalgh, C., Rodden, T., & Pycock, J. (2001). Collaborative virtual environments. *Communications of the ACM, 44*(7), 79-85.

Cauell, J., Bickmore, T., Campbell, L., & Vilhjálmsson, H. (2000). Designing embodied conversational agents. *Embodied conversational agents*, 29-63.

Chang, Y.-W., Hsieh, C.-J., Chang, K.-W., Ringgaard, M., & Lin, C.-J. (2010). Training and testing low-degree polynomial data mappings via linear SVM. *Journal of Machine Learning Research, 11*(Apr), 1471-1490.

Cheng, Y., Chiang, H.-C., Ye, J., & Cheng, L.-h. (2010). Enhancing empathy instruction using a collaborative virtual learning environment for children with autistic spectrum conditions. *Computers & Education, 55*(4), 1449-1458.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences 2nd edn. Erlbaum Associates, Hillsdale.

Constantino, J. N., & Gruber, C. P. (2002). The social responsiveness scale. *Los Angeles: Western Psychological Services*.

Cureton, E. E. (1956). Rank-biserial correlation. *Psychometrika, 21*(3), 287-290.

Curtis, D. D., & Lawson, M. J. (2001). Exploring collaborative online learning. *Journal of Asynchronous learning networks, 5*(1), 21-34.

Desmarais, M. C., & Baker, R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction, 22*(1-2), 9-38.

Eugenio, B. D., & Glass, M. (2004). The kappa statistic: A second look. *Computational linguistics, 30*(1), 95-101.

Gogoulou, A., Gouli, E., & Grigoriadou, M. (2008). Adapting and personalizing the communication in a synchronous communication tool. *Journal of Computer Assisted Learning, 24*(3), 203-216.

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics, 11*(1), 1-21.

Hourcade, J. P., Williams, S. R., Miller, E. A., Huebner, K. E., & Liang, L. J. Evaluation of tablet apps to encourage social interaction in children with autism spectrum disorders. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2013* (pp. 3197-3206): ACM

Jain, Y. K., & Bhandare, S. K. (2011). Min max normalization based data perturbation method for privacy protection. *International Journal of Computer & Communication Technology, 2*(8), 45-50.

Klakow, D., & Peters, J. (2002). Testing the correlation of word error rate and perplexity. *Speech Communication, 38*(1-2), 19-28.

Kopp, S., Gesellensetter, L., Krämer, N. C., & Wachsmuth, I. A conversational agent as museum guide–design and evaluation of a real-world application. In *International Workshop on Intelligent Virtual Agents, 2005* (pp. 329-343): Springer

Krishnaiah, P. R. (1980). *Handbook of statistics* (Vol. 31): Motilal Banarsidass Publishe.

Kumar, R., Rosé, C. P., Wang, Y.-C., Joshi, M., & Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. *Frontiers in artificial intelligence and applications, 158*, 383.

Leman, P. J. (2015). How do groups work? Age differences in performance and the social outcomes of peer collaboration. *Cognitive science, 39*(4), 804-820.

Linton, F., Goodman, B., Gaimari, R., Zarrella, J., & Ross, H. Student modeling for an intelligent agent in a collaborative learning environment. In *International Conference on User Modeling, 2003* (pp. 342-351): Springer

Mathiowetz, V., Weber, K., Volland, G., & Kashman, N. (1984). Reliability and validity of grip and pinch strength evaluations. *Journal of Hand Surgery, 9*(2), 222-226.

McManus, M. M., & Aiken, R. M. (1995). Monitoring computer-based collaborative problem solving. *Journal of Interactive Learning Research, 6*(4), 307.

Millen, L., Hawkins, T., Cobb, S., Zancanaro, M., Glover, T., Weiss, P. L., et al. Collaborative technologies for children with autism. In *Proceedings of the 10th International Conference on Interaction Design and Children, 2011* (pp. 246-249): ACM

Nabeth, T., Razmerita, L., Angehrn, A., & Roda, C. (2005). InCA: a cognitive multi-agents architecture for designing intelligent & adaptive learning systems. *Computer Science and Information Systems, 2*(2), 99-114.

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.

Pruette, J. R. (2013). Autism Diagnostic Observation Schedule-2 (ADOS-2).

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*: Western Psychological Services.

Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010). Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning, 5*(1), 43-102.

Schmidt, M., Laffey, J., & Stichter, J. Virtual social competence instruction for individuals with autism spectrum disorders: Beyond the single-user experience. In *Proceedings of CSCL, 2011* (pp. 816-820)

Schmidt, M., Laffey, J. M., Schmidt, C. T., Wang, X., & Stichter, J. (2012). Developing methods for understanding social behavior in a 3D virtual learning environment. *Computers in Human Behavior, 28*(2), 405-413.

Srinivasan, S., & Petkovic, D. Phonetic confusion matrix based spoken document retrieval. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, 2000* (pp. 81-87): ACM

Van Rosmalen, P., Brouns, F., Tattersall, C., Vogten, H., Bruggen, J., Sloep, P., et al. (2005). Towards an open framework for adaptive, agent-supported e-learning. *International Journal of Continuing Engineering Education and Life Long Learning, 15*(3-6), 261-275.

Vieira, A. C., Teixeira, L., Timóteo, A., Tedesco, P., & Barros, F. Analyzing on-line collaborative dialogues: The oxentchê–chat. In *International Conference on Intelligent Tutoring Systems, 2004* (pp. 315-324): Springer

Walker, E., Rummel, N., & Koedinger, K. R. (2014). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence in Education, 24*(1), 33-61.

Wallace, S., Parsons, S., & Bailey, A. (2015). Self-reported sense of presence and responses to social stimuli by adolescents with ASD in a collaborative virtual reality environment. *Journal of Intellectual and Developmental Disability*.

White, S. W., Keonig, K., & Scahill, L. (2007). Social skills development in children with autism spectrum disorders: A review of the intervention research. *Journal of autism and developmental disorders, 37*(10), 1858-1868.

Wilson, G. F., & Russell, C. A. (2007). Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. *Human factors: the journal of the human factors and ergonomics society, 49*(6), 1005-1018.

Zancanaro, M., Pianesi, F., Stock, O., Venuti, P., Cappelletti, A., Iandolo, G., et al. (2007). Children in the museum: an environment for collaborative storytelling. In *PEACH-Intelligent Interfaces for Museum Visits* (pp. 165-184): Springer.

CHAPTER VI. MULTIMODAL FUSION FOR COGNITIVE LOAD MEASUREMENTS

6.1. Abstract

In this chapter, a novel virtual reality (VR)-based driving system was introduced to teach driving skills to adolescents with ASD. This driving system is capable of gathering eye gaze, electroencephalography, and peripheral physiology data in addition to driving performance data. The objective of the current work is to fuse multimodal information to measure cognitive load during driving such that driving tasks can be individualized for optimal skill learning. Individualization of ASD intervention is an important criterion due to the spectrum nature of the disorder. Twenty adolescents with ASD participated in our study and the data collected were used for systematic feature extraction and classification of cognitive loads based on five well-known machine learning methods. Subsequently, three information fusion schemes—feature level fusion, decision level fusion and hybrid level fusion—were explored. Results indicate that multimodal information fusion can be used to measure cognitive load with high accuracy. Such a mechanism is essential since it will allow individualization of driving skill training based on cognitive load, which will facilitate acceptance of this driving system for clinical use and eventual commercialization.

6.2. Introduction

We have developed a novel VR-based driving system aimed at training driving skills in adolescents with ASD (Wade et al. 2016 (In press)). Training efficiency may be improved by adjusting difficulty levels of the driving tasks based on users' cognitive load. A cognitively intelligent system, which can sense, analyze and respond to a user's cognitive state has the potential to improve learning efficiency (Novak et al. 2012). For example, Koenig et al. implemented a cognitively intelligent system to maximize the training efficiency in their rehabilitation environments (Koenig et al. 2011). We will develop the VR-based driving system into a cognitively intelligent system because cognitive load appears to be more appropriate in the context of driving and is commonly used in driving related applications (Yannakakis

and Togelius 2011). This chapter explores fusion of multimodal information from a novel VR-based driving system for cognitive load measurement, which is a necessary step before building a cognitively intelligent VR-based driving system.

6.2.1. Background

Cognitive load is a multidimensional construct representing the working load that is imposed on a learner's cognitive system when performing a particular task (Fred GWC Paas and Van Merriënboer 1994). Cognitive load is believed to be a crucial factor in learning of complex tasks (F. Paas et al. 2003), such as driving tasks. The capacity of working memory is limited and it varies from person to person. If a learning task requires too little or too much cognitive capacity, learning may be impeded (De Jong 2010). Therefore it is important to design learning tasks that provide an appropriate level of cognitive load, which is neither too high nor too low (Schoor et al. 2012).

Cognitive load theory is concerned with efficient usage of people's limited working memory to acquire knowledge and skills. There are different types of cognitive load, such as intrinsic load and extraneous load (Sweller 2010). Intrinsic load reflects the natural complexity of learning information and the expertise of a learner. Extraneous load is related to the design of instructions (F. Paas et al. 2003). When task difficulty exceeds a learner's expertise, additional extraneous load is generated and the required cognitive load exceeds the learner's working memory capacity. When the learner's expertise exceeds the task difficulty, the learner wastes time and energy to solve tasks that are too simple and therefore will not benefit from learning. Thus the task difficulty level should match a learner's expertise in order to enable effective learning (Schnotz and Kürschner 2007).

Compared to TD individuals, working memory of individuals with ASD may be different (Rajendran and Mitchell 2007). Individuals with ASD performed significantly worse than TD individuals on tasks related to working memory (Bennetto et al. 1996). Remington et al. reported altered performance of individuals with ASD under different levels of cognitive load (Remington et al. 2009). Individuals with ASD also have difficulty in understanding the mental states of their own and others (Rajendran and

Mitchell 2007). Therefore, a targeted system that can automatically measure cognitive load of individuals with ASD and then optimize their cognitive load may have the potential to improve their learning efficiency (Ozonoff and Strayer 2001).

6.2.2. Related research

Real-time measurement of cognitive load in individuals with ASD is critical for a cognitively intelligent system. There are three general ways to measure cognitive load (Meshkati et al. 1995): subjective scales, performance-based measures and physiology-based measures. Subjective scales are inappropriate in a cognitively intelligent system for ASD intervention because: 1) individuals with ASD may have difficulty in accurately reporting their own cognitive load (Rajendran and Mitchell 2007), and 2) subjective scales are not real-time measures. We therefore explored measuring real-time cognitive load using information from eye gaze, electroencephalography (EEG), and peripheral physiology modalities along with a task performance modality.

Each of the above-mentioned modalities has been studied with regards to cognitive load measurement. It has been found that eye gaze signals are reflective of a user's cognitive state (Pomplun and Sunkara 2003). Pupil dilation is known to quickly respond to changes in a person's cognitive workload (Pomplun and Sunkara 2003). EEG signals are sensitive and reliable for continuous memory load measurement (Gevins et al. 1998). Alpha and theta wavebands of EEG are correlated with task difficulty (Gevins and Smith 2000). Peripheral physiological signals are also important components of cognitive load measurement (Mehler et al. 2009). Electrocardiogram (ECG), respiration (RSP), and HR were demonstrated to be sensitive to cognitive load in (Reimer et al. 2013; Novak et al. 2012). Performance-based measurement is a typical way to measure cognitive load (F. Paas et al. 2003). In terms of driving studies, performance metrics, such as steering wheel movements, lane-keeping behavior, speed control, and time-to-line crossing, have been found to be related to cognitive load (Son and Park 2011).

In order to classify cognitive load using observed information, several well-known machine learning algorithms and different parameter values of these algorithms have been evaluated in TD populations.

Hussain et al. tested k-nearest neighbor (KNN) with different k values in measuring cognitive load using face, physiology, and task performance data (M. S. Hussain et al. 2013). Different kernel functions of support vector machine (SVM), including linear kernel (M. S. Hussain et al. 2013) and Gaussian kernel (Son et al. 2013), were used in cognitive load measurement. Novak et al. analyzed linear discriminant analysis (LDA), diagonal LDA, and stepwise LDA to classify cognitive load (Novak et al. 2011). Lin et al. explored backpropagation and radial basis functions to build artificial neural networks (ANN) in cognitive load measurement (Lin et al. 2005). One of the most important parameters of building a decision tree is the splitting criterion (Narsky and Porter 2013). Hussain et al. selected cross-entropy as the splitting criterion to build decision trees for cognitive load measurement (M. S. Hussain et al. 2013). However, the studies using machine learning algorithms in measuring cognitive load of individuals with ASD are limited. Lagun et al. showed that SVM can achieve higher classification accuracy than naïve Bayes and logistic regression when measuring cognitive load of individuals with ASD (Lagun et al. 2011).

Fusing multimodal information to measure cognitive load has been explored in different applications. Novak et al. fused physiological and performance information for upper extremity rehabilitation (Novak et al. 2011). Steichen et al. used eye gaze together with performance information for cognitive load measurement in visualization systems (Steichen et al. 2014). Son et al. estimated users' cognitive workload using two spoken tasks by integrating performance and eye gaze information (Son and Park 2011). However, there is no study to our knowledge that has systematically studied fusing multimodal information to measure cognitive load of individuals with ASD during VR-based driving.

6.2.3. Current work

This work fuses multimodal information collected from a novel VR-based driving system for cognitive load measurement, which is a necessary step before building a cognitively intelligent VR-based driving system. We hypothesize that multimodal information can lead to a more accurate cognitive load measurement than single modality-based measurement approaches. This hypothesis is tested by

comparing single modality information to multimodal information in cognitive load measurement with multiple well known machine learning algorithms using data collected during VR-based driving in adolescents with ASD.

The key contribution of the current work is to design a cognitive load measurement technique for VR-based driving such that the driving difficulty can be adjusted for each individual based on their cognitive load, which will likely enhance learning. The ground truth of cognitive load used in this work is based on perceived task difficulty as experienced by the individuals with ASD and is rated by an experienced clinically trained rater. This ground truth, as we have shown, correlates well with the driving performance of users, and thus provides a method to measure cognitive load that overcomes the difficulty associated with self-rating, which is problematic for individuals with ASD. Thus this work contributes in the following aspects: 1) to analyze eye gaze, EEG, peripheral physiological and performance data in the context of VR-based driving, which is designed to provide a safe and flexible environment to teach driving skills to adolescents with ASD who often have deficits in this regard; 2) to extract useful features from these data that can be used to measure their cognitive load; and 3) to apply several machine learning algorithms for measuring cognitive load of a user as well as explore how multimodal information can be fused at different levels to yield highly accurate cognitive load measurement.

The work is organized as follows. Section 6.3 describes our novel VR-based driving system, including system design and experimental setup. Section 6.4 lists the features extracted from four modalities. Section 6.5 presents the classification algorithms as well as three data fusion strategies for cognitive load measurement. The results are provided in Section 6.6 followed by a discussion in Section 6.7. Finally conclusions of the presented work and future research plans are discussed in Section 6.8.

6.3. VR-Based Driving System

6.3.1. System design

A VR-based driving system was designed to train and improve the driving skills of adolescents with ASD. The three primary components of the VR-based driving system were: a driving simulator, a data capture module and a rating module, as shown in Fig. 25.

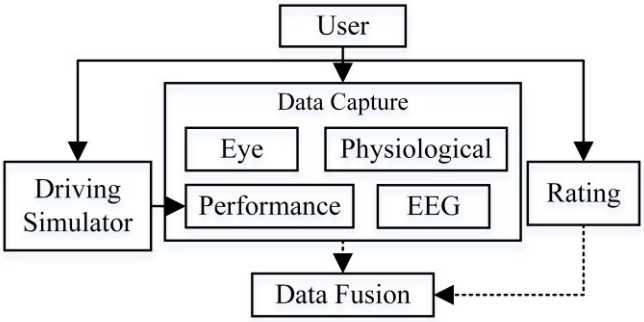

Fig. 25 The framework of VR-based driving system

Fig. 26 shows the driving simulator. A Logitech G27 steering wheel controller was used to control a virtual agent vehicle in the virtual driving environment. Models in the virtual driving environment, such as traffic lights, stop signs, and vehicles were developed with the modeling tools ESRI CityEngine (www.esri.com/cityengine) and Autodesk Maya (www.autodesk.com/maya). The game development platform Unity3D (www.unity3d.com) was used to implement the system logic. A total of six different difficulty levels, each level consisting of three driving assignments, were developed for the VR-based driving system.

Fig. 26 The driving simulator of the VR-based driving system

These difficulty levels were tested and validated in our previous works (Wade et al. 2016 (In press)). Control parameters (Table 33), such as speed of vehicles, responsiveness of the agent vehicles' brake and accelerator, and weather conditions, were manipulated to produce a range of difficulties. Table 34 shows the values of these control parameters used in each designed difficulty level.

Table 33 The Control Parameters of Difficulty Level

Label	Description of parameter	Domain
A_s	Speed of autonomous vehicles	$A_s \in [0.85, 1.75]$
A_a	Aggressiveness of autonomous vehicles	$A_a \in [1, 1.5]$
H_s	Traffic light alert sound.	$H_s \in$ {Enabled,Disabled}
R_b	Responsiveness of the brake pedal.	$R_b \in [0.35, 1]$
R_a	Responsiveness of the accelerator pedal.	$R_a \in [1, 1.5]$
R_s	Responsiveness of the steering wheel.	$R_s \in [1, 3.75]$
W	Weather condition.	$W \in$ {Sunny, Overcast,Rainy}
L	Intensity of light in the environment.	$L \in [0.01, 0.5]$
N_v	Number of vehicles at intersections.	$N_v \in \{1, 2, ..., 5\}$
S_d	Duration of time to permit driving on sidewalk.	$S_d \in [0.6, 4]$

Table 34 The Configuration of the Designed Difficult Level

Level	As	Aa	Hs	Rb	Ra	Rs	W	L	Nv	Sd
1	0.85	1	Enabled	1	1	1	sunny	0.5	0 to 1	4
2	1	1	Disabled	1	1	1	sunny	0.466	1 to 2	3.35
3	1.35	1	Disabled	1	1	1	overcast	0.409	2 to 3	2.66
4	1.35	1	Disabled	1	1	1	sunny	0.329	2 to 3	1.97
5	1.35	1.35	Disabled	0.675	1.25	2.375	sunny	0.226	3 to 5	1.29
6	1.75	1.5	Disabled	0.35	1.5	3.75	rainy	0.01	3 to 5	0.6

The data capture module recorded a user's multimodal information while the user was engaged in driving. A Tobii X120 remote eye tracker (www.tobii.com) logged the eye gaze data at 120 Hz. A Biopac MP150 (www.biopac.com) physiological data acquisition system wirelessly sampled multiple peripheral physiological signals, including ECG, electromyography (EMG), RSP, SKT, photoplethysmogram (PPG), and galvanic skin response (GSR). The PPG and GSR signals were measured from toes instead of fingers in order to reduce the motion artifact from driving. The SKT signal was collected from the upper arm. An Emotiv EPOC wireless EEG headset (www.emotiv.com) recorded 14-channel EEG signals. Metrics of the user's performance was recorded within the virtual driving environment.

We designed a rating mechanism for a rater to observe and rate a user's affective and cognitive state in real time. A live video with sound, recording of the user's frontal face and the virtual driving environment, was displayed for the rater. The rater could also view the entire experimental environment via a one-way mirror from an adjacent room. The computer used by the rater was connected to the driving simulator and the data capture module via a local area network (LAN). The data of each system component were labeled according to timestamps of the driving simulator component in order to facilitate offline synchronization.

6.3.2. Experimental setup

A total of 20 adolescents with ASD, from 13 to 18 years old, were involved in a series of six experimental sessions. The participants were recruited through an existing university based clinical research registry. Although the study was open to adolescents from both genders, the majority (19 out of 20) were male participants. ASD is much more common in males than in females (Werling and Geschwind 2013) and we were not able to recruit more female participants. All participants had a clinical diagnosis of ASD from a licensed clinical psychologist. The Social Responsiveness Scale, second edition (SRS-2) was completed for each participant by his/her parent to quantify the severity of his/her ASD symptoms (Kim and André 2008). This study was approved by the Vanderbilt University Institutional Review Board (IRB). Table 35 shows detailed participants' information.

Table 35 The Participants' Information

Gender (%male)	Age (year)	SRS-2 total raw score	SRS-2 score
95%	15.29(1.66)	97.85(28.35)	75.45(10.23)

Each of the participants completed six sessions on different days. Each session lasted approximately one hour. Fig. 27 shows the experimental protocol of a session. The blocks with dashed lines represent experimental steps that are only parts of the first session. At the beginning of the first session, informed consent was obtained. A video tutorial regarding the VR-based driving system was then shown to a participant in the first session. Three researchers set up peripheral physiological and EEG sensors, and calibrated an eye tracker in the sensor application step. Before data recording, all signals were checked by

the researchers to make sure all the sensors were placed correctly. Then, baseline data were collected for three minutes for the peripheral physiological and EEG signals in a silent environment. In the first session, after the baseline recording step, the participant took part in driving tasks in a free-form mode. Following this step, three pre-selected driving assignments were carried out. During the driving assignments, the researchers monitored the peripheral physiological and EEG signals in real time to ensure the quality of the recorded data.

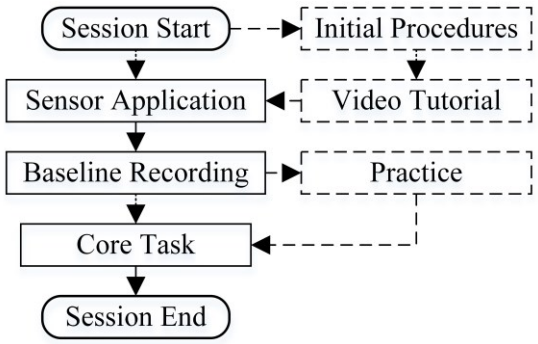

Fig. 27 The experimental protocol of a session

The first and the last sessions acted as pre- and post-tests and included the same three driving assignments (i.e., one easy driving assignment and two difficult driving assignments). The pre- and post-tests were included in order to evaluate the system in improving a participant' diving skills. However, we do not consider performance improvement from the pre-test to the pose-test in this work. Each of the other four sessions were composed of three driving assignments from the same difficulty level, with the driving difficulty increasing from the second to the fifth sessions.

During the experiment, a rater rated a participant's affective and cognitive states in real time using the rating mechanism described in Section 6.3.1. The rater had extensive experience working with individuals with ASD at the Treatment and Research Institute for Autism Spectrum Disorders (TRIAD) at Vanderbilt University. The rater had been trained to utilize a rating system across a series of other works regarding human-computer interaction in ASD populations (C. Liu et al. 2009). She was directly supervised by licensed clinical psychologists who specialized in ASD diagnosis and treatment. Five categories of rating were collected: perceived task difficulty level, engagement, enjoyment, boredom, and frustration.

However, only the rating of perceived task difficulty is considered in this work. The rater rated the perceived task difficulty experienced by the participant in a continuous interval from 0 to 9 using 5 as the threshold. Specifically, a task was rated with a value higher than 5 if it was perceived to be hard with larger value indicating higher perceived task difficulty, and vice versa. The continuous rating of perceived task difficulty was later mapped into binary classes offline using 5 as the threshold. That is, if a rating of perceived task difficulty had a value less than five, it was mapped into the low cognitive load class; otherwise, it belonged to the high cognitive load class.

We did not use the designed task difficulty as ground-truth for cognitive load considering that the cognitive load caused by the same task may vary: 1) from person to person, and 2) at different times for the same person (Schnotz and Kürschner 2007). We therefore utilized the rating of perceived task difficulty by a trained rater for the ground truth of cognitive load. It was assumed that a high rating of perceived task difficulty was indicative of a high cognitive load experienced by the participant (Fred G Paas 1992).

6.4. Feature extraction

6.4.1 Eye gaze features

The eye tracker signals, recorded by the Tobii X120 eye tracker, were preprocessed in order to remove invalid data and reduce noise. If the time duration of continuous lost data was larger than 1000 ms, the lost data were removed. This long duration lost data were primarily attributed to the movement of a participant's head beyond the eye tracker's detection range. If the time duration of continuous lost data was less than 75ms, the lost data were filled in with valid data using a linear interpolation method (Olsen 2012). 75ms was selected as the threshold because it is the minimum closure duration of a blink. Any lost eye gaze data with a duration less than 75ms was deemed to be due to noise. The noise in the eye gaze data was then reduced with a median filter.

After preprocessing, we extracted 4 basic eye gaze features guided by previous literature (Lahiri et al. 2011; Pomplun and Sunkara 2003): blink, pupil diameter, fixation and saccade (Table 36). We then

extracted 10 secondary eye gaze features from the basic features, i.e. blink rate, fixation rate, Mean and Standard Deviation (M and SD) of blink duration, M and SD of pupil diameter, M and SD of fixation duration, and M and SD of saccade duration.

Table 36 Basic Eye Gaze Feature

Basic features	Definition
Blink	A rapid closing of eye with closure duration between 75 ms to 400 ms
Pupil diameter	The pupil diameter, unit in mm
Fixation	The eye gaze maintains in one point with a very slow eye movement
Saccade	A quick eye movement between two fixations.

6.4.2 EEG features

We recorded EEG signals using the Emotiv EPOC neuroheadset. EEG signals were collected at 128 Hz from 14 channels at locations AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 as defined by the international 10-20 system (Klem et al. 1999). The reference sensors were placed at locations P3 and P4. The recorded EEG signals had bandwidth from 0.2 Hz to 45 Hz, covering five frequency bands, which are delta (frequency<4Hz), theta (4Hz<frequency<8Hz), alpha (8Hz<frequency<13Hz), beta (13Hz<frequency<30Hz), and gamma (frequency> 30 Hz). The theta, alpha, beta, and gamma frequency band activities have been reported to be sensitive for measuring cognitive load for ASD populations (Lushchekina et al. 2013). The delta band was less informative and was susceptible to movement artifacts during driving. Therefore, we excluded the delta band from feature extraction in this work.

The raw EEG signals were preprocessed by removing the outliers, which were defined as the change between two adjacent data points $>50\,\mu v$. Then, a low pass filter and a high pass filter were used to remove the noise with frequency larger than 45Hz and less than 0.2Hz. After filtering, data were chopped into 1s epoch and those with poor contact quality were rejected. Eye blink, eye movement, and muscle movement artifacts were removed with an EOG-EMG artifact correction algorithm (De Clercq et al. 2006).

The power spectral density was the feature that best reflected the changes of the EEG activities and therefore was utilized in the previous literature for cognitive load measurement (Antonenko et al. 2010). In this work, power spectral density variables of theta, alpha, beta, and gamma bands, were extracted from the preprocessed signals in each channel, resulting in a total number of 56 features (14 channels × 4 wave bands = 56 features).

6.4.3. Peripheral physiological features

We used the Biopac MP150 and recorded ECG, EMG, RSP, SKT, PPG, and GSR signals with a 1000Hz sampling frequency. The EMG signals were recorded from Corrugator, Zygomaticus, and Trpezius muscles. These peripheral physiological signals were preprocessed offline with three steps: 1) outlier removal; 2) noise reduction with filters; and 3) subsampling. The details about how to analyze the peripheral physiological data can be found in (Sarkar 2002). In the first step, very small and very large outliers in each peripheral physiological signal were removed separately. Then, the noise was reduced using a low pass filter, a high pass filter, and a notch filter. The slowly changing signals, SKT, RSP, and GSR, were subsampled to reduce computation. The subsampling equation used with k=10 was: $x_{subsample}[n] = x_{initial}[kn]$. We identified 60 features from peripheral physiological signals as shown in Table 37.

Table 37 The Peripheral Physiological Features and Their Descriptions

Signal	Basic features	Description
PPG	Amplitude of peak values (M and SD)	The amplitude of the detected pulse
	Pulse Transit Time (M and SD)	The width of the detected pulse
GSR	Tonic activity level (M and SD)	Tonic level of electrical conductivity of skin
	Slope of tonic activity	The change of the tonic per second
	Amplitude phasic activity (M and SD)	The amplitude of detected skin conductance response (SCR) peak
	Rate of phasic activity	The number of the detected SCR peak per second
	Rise time (M and SD)	Temporal interval between SCR initiation and SCR peak
	Recovery time (M and SD)	Temporal interval between SCR peak and point of 50% recovery of SCR amplitude
EMG	EMG activity (M and SD)	One of the EMG signal
	Slope of activities	The slow change of one EMG signal per minute
	burst activities frequency	Number of EMG burst peak per minute
	The burst activities (M and SD)	The time duration of EMG burst peak
	Activity frequency (M and SD)	The frequency of one EMG signal
	Amplitude of burst activities	The amplitude of the detected EMG burst peak
RSP	Amplitude (M and SD)	The amplitude of the detected breath peak
	Subband spectral entropy	The spectral entropy in three subband 0.003-0.04Hz, 0.04-0.15Hz, and 0.15-0.4Hz
	Minimum and maximum difference	The difference between the minimum and the maximum amplitude of detected breath peak
	peak frequency	The number of the detected breath peak per minute
	Power spectrum density of low power	The power of low-frequency component (0.04-0.15Hz)
	Power spectrum density of high power	The power of high-frequency component (0.15-0.4Hz)
	The first order difference	The output of the first-order difference equation
	Poincare plot geometry SD1	The variance corresponding to short-term breathing rate variability.
	Poincare plot geometry SD2 (M and SD)	The variance corresponding to long-term breathing rate variability.
	Peak valley magnitude (M and SD)	The magnitude between the peak and the valley
	Respiratory rate (M and SD)	The number of breaths per minute
SKT	Temperature (M and SD)	Peripheral temperature united in degree centigrade
	Slope of temperature	The change of the temperature per second

6.4.4. Performance features

In the described work, participants' performance data were recorded through their driving behavior and task performance. The driving behavior included how a participant used the brake and accelerator during driving. The task performance indicated how well a participant completed a task, such as how many times

he/she failed in one assignment and the driving score he/she achieved during one driving assignment. All the performance features and their descriptions are listed in Table 38.

Table 38 Performance Features and Their Description

Features	Description
Brake (M and SD)	The level of using brake. A value between 0 and 1. 0 means no brake. 1 means full brake.
Accelerator (M and SD)	The level of using accelerator. A value between 0 and 1. 0 means no acceleration. 1 means full acceleration.
Failure times	The number of driving failures
Driving score	Number of points achieved during one assignment

6.5. Classification and Data Fusion Method

6.5.1 The classification algorithm

We applied five well-known classification algorithms: SVM (Bishop 2006), KNN (Aditya and Tibarewala 2012), decision tree (Safavian and Landgrebe 1990), discriminant analysis (Fisher 1936), and ANN (Hagan et al. 1996), to classify the cognitive load from recorded data. Because the accuracy of each machine learning algorithm depended on its key parameter (Bergstra and Bengio 2012), we tested each machine learning algorithm with a variety of parameter values. Table 39 summarizes the evaluated classification algorithms and specifies their parameter values used for cognitive load measurement in this work. Regarding SVM, the value of the C parameter was 1 and the size of the radial basis function was also 1, which are used in this work because they resulted in high accuracy in our previous work (C. Liu et al. 2009). In terms of ANN, the value of the calculated number of hidden neurons is given by $(N_f + N_c)/2$, where N_f is the input feature number and N_c is the output class number.

6.5.2 Data fusion methods

In general, multimodal information can be fused in different levels: feature level fusion, decision level fusion, and hybrid level fusion (Atrey et al. 2010). Feature level fusion is easy to use but is not robust if information of some modalities is lost. Decision level fusion is a more robust method that combines the sub-decision of each modality (Koelstra et al. 2012). The disadvantage of decision level fusion is its

failure to reflect the correlation between features of different modalities (E. S. Liu and Theodoropoulos 2014). Hybrid level fusion methods seek to combine the advantages of feature level fusion and decision level fusion (Atrey et al. 2010). However, it is not clear which level of fusion gives the highest accuracy in cognitive load measurement with eye gaze, EEG, peripheral physiological, and performance data in the VR-based driving system. We therefore compared these three fusion levels in fusing multimodal information in cognitive load measurement. The frameworks of the three level fusion techniques are shown in Fig. 28.

Table 39 The List of Classification Algorithms Used to Measure Cognitive Load

Classifier Index	Algorithm	Parameters and their values
1	SVM	Linear kernel
2		Quadratic kernel
3		Polynomial kernel of degree 3
4		Gaussian radial basis function kernel
5	KNN	Euclidean distance and k=1
6		Euclidean distance and k=3
7		Euclidean distance and k=5
8		Covariance distance and k=1
9		Covariance distance and k=3
10		Covariance distance and k=5
11		Cosine distance and k=1
12		Cosine distance and k=3
13		Cosine distance and k=5
14	Decision Tree	Gini's diversity index as split criterions
15		Deviance as split criterions
16		Twoing as split criterions
17	Discriminant analysis	Linear discriminant analysis
18		Quadratic discriminant analysis
19	ANN	Conjugate gradient backpropagation and with 10 hidden neurons
20		RPROP algorithm and with 10 hidden neurons
21		Marquardt algorithm and with 10 hidden neurons
22		Conjugate gradient backpropagation and with calculated number of hidden neurons
23		RPROP algorithm and with calculated number of hidden neurons
24		Marquardt algorithm and with calculated number of hidden neurons

Fig. 28 (a) shows the framework of feature level fusion. The input to the feature level fusion is a feature vector, which is composed of features from eye gaze modality (Eye), EEG modality (EEG), peripheral physiology modality (Phy), and performance modality (Per). In the preprocessing module, each feature of the feature vector is first normalized into a range from 0 to 1. Then, the dimension of the

feature vector is reduced with principal component analysis. A classifier takes the feature vector after preprocessing as input and outputs a level of cognitive load (CL).

Fig. 28 (b) shows the framework of decision level fusion. Each of the four modalities yields a feature vector. Each feature vector is preprocessed in a preprocessing module as discussed in feature level fusion. Because the dimensions of the feature vectors extracted from eye gaze and peripheral 153hysiology modalities are small, dimension reduction is not needed for these feature vectors. After preprocessing, each feature vector is input into a classifier, which outputs a level of cognitive load as a sub-decision. The fusion module calculates the final decision based on the weighted average of the four sub-decisions (2). The weighted average, y, is a function of a sub-decision vector, $\mathbf{D}$, and a weight vector, $\mathbf{W}$, (3). The elements of the sub-decision vector are four sub-decisions $\mathbf{D} = (d_1, d_2, d_3, d_4)$. Each sub-decision is an output of a binary classifier and therefore its value can be either 0 (meaning a low level of cognitive load) or 1 (meaning a high level of cognitive load). The elements of the weight vector are four weights, $\mathbf{W} = (w_1, w_2, w_3, w_4)$. Each weight is in the range $[0,1]$ and the sum of all four weights is 1.

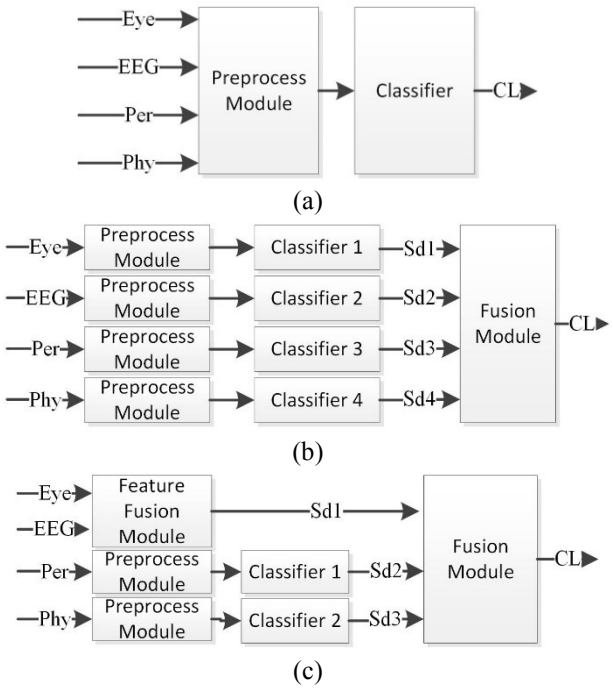

Fig. 28 (a) Feature level fusion framework; (b) decision level fusion framework; and (c) hybrid level fusion framework

$$d_{final} = \begin{cases} 0, & y<0.5 \\ 1, & y\geq 0.5 \end{cases} \quad (2)$$

$$y = \mathbf{W}\mathbf{D}^T = \sum_{i=1}^{4} w_i d_i \quad (3)$$

The final decision depends on the weight vector. The weight vector that produces the highest accuracy of decision level fusion is the optimal weight vector, which is usually found by exhaustive search. For example, Koelstra et al. incremented each weight of a two-dimensional weight vector from 0 to 1 by 0.01 in order to find an optimal weight vector for emotion recognition using EEG and peripheral physiological signals (Koelstra et al. 2012). However, the exhaustive search method is computationally expensive for decision level fusion with a high-dimensional weight vector. For example, a decision level fusion with a four-dimensional weight vector using exhaustive search with 0.01 step width needs to evaluate 10^6 weight vectors in order to find the optimal one. Because that the sub-decisions of our decision level fusion were binary data, the search space of weight vectors can be reduced. We present a new approach that allows finding an optimal weight vector from a small number of weight vectors and thereby reducing computational load significantly. We prove that the optimal weight vector can be found from the small number of selected weight vectors.

Lemma 1: A small number of weight vectors can yield the optimal one for a decision level fusion with four binary sub-decisions.

Proof: We define a small number of sets of weight vectors that cover the optimal one for the decision level fusion (Step 1). We prove that all weight vectors of each set yield the same final decision (Step 2).

Step 1: The universal set of weight vectors can be presented as (4).

$$U = \{(w_1, w_2, w_3, w_4) \in [0, 1]^4 | \sum_{i=1}^{4} w_i = 1\} \quad (4)$$

First, based on whether $w_{\max}$ (the maximum weight of a weight vector in U (6)) is greater than, equal to, or less than 0.5 , the universal set U can be partitioned into three disjoint subsets: O , P, and Q , respectively. If a weight vector is a member of the subset O as defined by (7), the final decision is determined by the sub-decision associated with the maximum weight of the weight vector. This condition is discussed separately in the results section as the single modality classification. If a weight vector is a member of the subset P as defined by (8), the decision level fusion produces a low accuracy because of

the possible boundary condition, $y = 0.5$. We, therefore, excluded the subset P for the decision level fusion in this work. The subset Q as defined by (8) is a collection of weight vectors, where the weights are each less than 0.5.

$$U = O \cup P \cup Q \tag{5}$$

$$w_{\max} = \max(w_1, w_2, w_3, w_4) \tag{6}$$

$$O = \{(w_1, w_2, w_3, w_4) \in U \mid w_{\max} > 0.5\} \tag{7}$$

$$P = \{(w_1, w_2, w_3, w_4) \in U \mid w_{\max} = 0.5\} \tag{8}$$

$$Q = \{(w_1, w_2, w_3, w_4) \in U \mid w_{\max} < 0.5\} \tag{9}$$

Second, based on whether $w_{\max} + w_{\min}$ (the sum of the maximum weight of a weight vector from (6) and the minimum weight of the weight vector from (11)) is greater than, equal to, or less than 0.5, set Q can be partitioned into three disjoint subsets as shown by (10). We excluded Q_C for the decision level fusion in this work. If a weight vector is a member of Q_C, the decision level fusion produces a low accuracy due to the possible boundary condition, $y = 0.5$.

$$Q = Q_A \cup Q_B \cup Q_C \tag{10}$$

$$w_{\min} = \min(w_1, w_2, w_3, w_4) \tag{11}$$

$$Q_A = \{(w_1, w_2, w_3, w_4) \in Q \mid w_{\max} + w_{\min} > 0.5\} \tag{12}$$

$$Q_B = \{(w_1, w_2, w_3, w_4) \in Q \mid w_{\max} + w_{\min} < 0.5\} \tag{13}$$

$$Q_C = \{(w_1, w_2, w_3, w_4) \in Q \mid w_{\max} + w_{\min} = 0.5\} \tag{14}$$

Third, the set Q_A can be further partitioned into four subsets according to the index of the maximum weight of a weight vector in Q_A, i.e. $Q_{A1} = \{(w_1, w_2, w_3, w_4) \in Q_A \mid w_1 = w_{\max}\}$,…, and $Q_{A4} = \{(w_1, w_2, w_3, w_4) \in Q_A \mid w_4 = w_{\max}\}$. The maximum weight of the weight vector in Q_A is unique. This can be shown by the fact that a weight vector with more than one maximum weights will result in an invalid sum of the vector's elements: $w_1 + w_2 + w_3 + w_4 \geq 2w_{\max} + 2w_{\min} > 1$. Therefore, these subsets of Q_A are disjoint sets.

Fourth, the set Q_B can be further partitioned into four subsets according to the index of the minimum weight of a weight vector in Q_B, i.e. $Q_{B1} = \{(w_1, w_2, w_3, w_4) \in Q_B \mid w_1 = w_{\min}\}$,…, and $Q_{B4} = \{(w_1, w_2, w_3, w_4) \in Q_B \mid w_4 = w_{\min}\}$. It is easy to prove that the subsets of Q_B are disjoint sets. These eight disjoint sets, $Q_{A1}, Q_{A2}, \ldots, Q_{B4}$, were considered in this work for decision level fusion.

Step 2: We prove that, within each of these eight subsets, the final decision of the decision level fusion is independent of the choice of the weight vector. We prove this assertion using the subset, Q_{A1} (15) as an example.

$$Q_{A1} = \{(w_1, w_2, w_3, w_4) \in Q \mid w_1 = w_{\max}, w_{\max} + w_{\min} > 0.5\} \quad (15)$$

The value of an element of a sub-decision vector is 0 or 1 and, therefore, there are a total of $2^4 = 16$ sub-decision vectors. Any sub-decision vector belongs to one of the four cases shown below by (16) to (19).

$$\text{Case 1: } d_1 = 1,\ \exists d_k = 1,\ k = 2,3,4 \quad (16)$$

$$\text{Case 2: } d_1 = 1,\ \forall d_k = 0,\ k = 2,3,4 \quad (17)$$

$$\text{Case 3: } d_1 = 0,\ \exists d_k = 0,\ k = 2,3,4 \quad (18)$$

$$\text{Case 4: } d_1 = 0,\ \forall d_k = 1,\ k = 2,3,4 \quad (19)$$

The final decision associated with weight vectors in Q_{A1} is shown in Table 40. As can be seen, if a weight vector is in Q_{A1}, the final decision is dependent on the sub-decision vector, but is independent of the weight vector. Therefore, a weight vector of Q_{A1} can represent all its weight vectors. It follows, then, that we can prove the assertion for any of the 8 subsets.

Table 40 The Values of Final Decision when Sub-decision in Different Cases

W	D	y	d_{final}
Q_{A1}	Case 1	$\sum_{i=1}^{4} w_i d_i \geq w_1 + w_{\min} > 0.5$	1
Q_{A1}	Case 2	$\sum_{i=1}^{4} w_i d_i = w_1 < 0.5$	0
Q_{A1}	Case 3	$\sum_{i=1}^{4} w_i d_i \leq (1 - (w_1 + w_{\min})) < 0.5$	0
Q_{A1}	Case 4	$\sum_{i=1}^{4} w_i d_i = 1 - w_1 > 0.5$	1

In conclusion, we defined a small number of subsets that cover the optimal weight vector for the decision level fusion. We proved that all weight vectors of a subset yielded the same final decision. Therefore, we can find the optimal weight vector by, 1) randomly selecting a weight vector from each of the eight subsets; and 2) computing and comparing accuracies of the decision level fusion with these eight weight vectors. The weight vector that yields the highest accuracy is the optimal one. Thus a small number of weight vectors can yield the optimal one for a decision level fusion with four binary sub-decisions and hence proves the lemma 1.

Fig. 28 I shows one instance of hybrid level fusion. Hybrid level fusion combines the processes of the feature level fusion and decision level fusion. The feature fusion module takes multimodal features (Eye

and EEG in Fig. 28 I) as input and outputs a level of cognitive load as a sub-decision. Each of other sub-decisions is calculated by inputting the feature vector of one modality into a classifier. The final decision of hybrid level fusion is the weighted average of all sub-decisions.

We have calculated results of the hybrid level fusion with two sub-decisions and with three sub-decisions. Hybrid level fusion with one sub-decision is equivalent to feature level fusion; while hybrid level fusion with four sub-decisions is equivalent to decision level fusion. All the possible combinations of different modalities' features were tested for the feature fusion module of the hybrid level fusion, which is listed in Section 6.6.3.

6.6. Results

Each participant completed six experimental sessions. Each session included 3 driving assignments. A binary cognitive load label (i.e., 0 or 1) was assigned to each driving assignment. Each driving assignment yielded one data sample. A total of 360 data samples were extracted (20 participants × 6 senssions × 3 assignments = 360 samples). However, because of data loss during the experiments, mostly due to the movement of participants, 74 bad data samples were removed after preprocessing. Ultimately, 286 data samples were included for the data analysis.

K-fold cross validation was selected to evaluate classification results. Usually, 5- to 10-fold cross validation is used in the literature to compute classification accuracy. In this work, 5-fold cross validation was selected so that enough test data were included for validation. We ran the 5-fold cross validation 10 times and averaged their results as the final accuracy in order to make the result more robust.

6.6.1. Analysis of rating of perceived task difficulty

Fig. 29 depicts a histogram of ratings of perceived task difficulty for data analysis (M = 5.28, SD = 1.39). As can be seen, a large portion of the ratings of perceived task difficulty lie around 5, which means a majority of the driving tasks were perceived at medium difficulty level by the participants. This distribution fits our goal of training driving skills of adolescents with ASD because very easy or very hard tasks are not conducive to train driving skills. 57.34% of all the assignments were labeled as high level

cognitive load, while 42.66% data samples have low cognitive load labels. For the almost balanced data, an accuracy was used to evaluate performance of classification models.

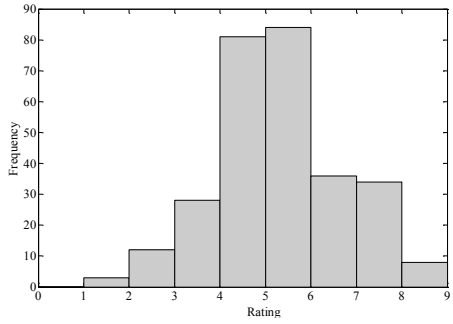

Fig. 29 The histogram of the rating of perceived task difficulty

Performance is an implicit estimation of cognitive load (Miller 2001). Performance features, such as reaction time (S. Hussain et al. 2011) and success frequency (Wu et al. 2010), were previously utilized to evaluate the ground truth of cognitive load. In this work, the correlation between ratings of perceived task difficulty and driving scores (a driving score is a performance feature, as shown in Table 38, and represents the success frequency in an assignment) was tested. The statistical analysis method, Spearman rank correlation, was selected because the driving score was an ordinal variable (Mukaka 2012). There was a strong negative correlation between the driving score and the rating of perceived task difficulty, $\rho(284)$=-0.62 , $p < 0.01$. No correlation between the driving score and the designed difficulty level was found, $\rho(284)$=0.06 , p=0.93 , from the experimental data. Because performance is an indicator of cognitive load (Miller 2001), these correlation results support that the rating of perceived task difficulty was a more reliable ground truth for cognitive load as compared to the designed difficulty level.

6.6.2. Feature level fusion and single modality classification

The first hypothesis of this work is that by combining multimodal information, the accuracy of cognitive load measurement will increase. The hypothesis is tested by comparing multimodal information to each single modality information in cognitive load measurement with several classifiers. The choice of a classifier is data dependent (Bishop 2006). Thus, a classifier may be insufficient to show the impact of different datasets in cognitive load measurement. We selected several commonly used classifiers, shown

in Table 39, for cognitive load measurement. All classifiers were used in feature level fusion and each single modality classification. Their accuracies are shown in Table 41 for the purpose of comparison. The best accuracy of feature level fusion and the best accuracy of each single modality classification are shown in bold font type. The average accuracy of feature level fusion and the average accuracy of each single modality classification are shown at the bottom of the table. The best accuracy of feature level fusion, 84.43%, is higher than the best accuracy of each single modality classification. On average, feature level fusion also achieved a higher accuracy compared to each single modality classification. The accuracy of feature level fusion was statistically significantly higher than the accuracy of each single modality classification, i.e. the accuracy of the eye gaze based classification ($Z = -4.88$, $p < .001$), the accuracy of the EEG based classification ($Z = -1.97$, $p < .05$), the accuracy of the peripheral physiological information based classification ($Z = -2.96$, $p < .05$), and the accuracy of the performance based classification ($Z = -4.61$, $p < .001$). These results suggest that combining multimodal information has the ability to increase the accuracy of cognitive load measurement.

6.6.3. Decision level fusion and hybrid level fusion

The final decision of decision level fusion is a weighted average of four sub-decisions as discussed in Section 6.5.2. All classifiers listed in Table 39 were used for each of the four sub-decisions in the decision level fusion. Each possible weight set (described in Section 6.5.2) was tested for the weighted average in decision level fusion. The highest observed accuracy of decision level fusion was 81.48%.

There are two types of hybrid level fusion for our data: hybrid level fusion with three sub-decisions and hybrid level fusion with two sub-decisions as discussed in Section 6.5.2. The best accuracies for these two types of hybrid level fusion are shown in Table 42 and Table 43, respectively. In both tables, the column *Classifier* indicates the Classifier index in Table 39 that gives the best accuracy when classifying cognitive load using the corresponding features. The highest accuracy of hybrid level fusion was 83.42%.

Table 41 Accuracies of All Algorithms/Parameters (%)

Classifier Index	Eye	EEG	Phy	Per	Fusion
1	58.45	64.10	65.52	69.77	73.66
2	60.15	69.08	67.04	68.66	73.65
3	63.92	72.93	71.21	61.45	78.13
4	73.16	78.18	70.58	65.44	81.53
5	72.83	79.33	78.77	64.11	82.80
6	67.99	76.64	73.00	60.46	78.48
7	62.17	76.46	73.36	59.47	77.94
8	70.94	79.96	79.31	63.75	84.43
9	68.73	76.35	75.76	62.79	79.27
10	61.89	77.29	74.92	63.08	79.34
11	71.27	79.62	78.71	65.19	81.77
12	66.16	77.63	73.90	61.63	79.56
13	63.60	77.52	74.98	60.78	78.44
14	59.17	63.50	56.97	68.39	64.15
15	59.91	62.15	57.99	62.50	62.70
16	58.33	63.62	57.19	61.34	62.21
17	59.79	70.06	62.77	70.53	74.32
18	68.72	63.52	73.86	67.37	80.89
19	66.15	70.71	65.18	59.08	72.21
20	54.47	70.16	66.68	66.29	75.26
21	53.50	63.66	57.62	69.16	67.18
22	64.19	72.18	60.14	66.64	69.79
23	55.53	67.12	58.56	70.68	73.24
24	58.09	65.64	62.08	73.72	69.17
AVG	63.30	71.56	68.17	65.09	75.01

Table 42 Accuracies of Hybrid Level Fusion with Three Sub-decisions

Sub-decision 1		Sub-decision 2		Sub-decision 3		Accuracy
Features	Classifier	Features	Classifier	Features	Classifier	
performance	24	physiological	8	EEG & Eye gaze	4	81.35%
performance	24	physiological & EEG	8	Eye gaze	4	80.89%
performance & EEG	18	Physiological	8	Eye gaze	4	73.84%
performance	24	EEG	8	physiological & Eye gaze	8	80.57%
performance & physiological	24	EEG	8	Eye gaze	4	79.10%
performance & Eye gaze	8/4	physiological	8	EEG	8	79.46%

Table 43 Accuracies of Hybrid Level Fusion with Two Sub-decisions

Sub-decision 1		Sub-decision 2		Accuracy
Features	Classifier	Features	Classifier	
EEG	8	performance & Eye gaze & physiological	8	83.00%
Eye gaze	4	performance & EEG & physiological	8	83.42%
physiological	8	performance & Eye gaze & EEG	8	81.52%
performance	24	physiological & EEG & Eye gaze	8	82.86%

6.7. Discussion

6.7.1 Feature level fusion and single modality classification

We found that feature level fusion performed better than all single modality classifications in cognitive load measurement indicated by statistical tests results, their best accuracies, and average accuracies. There are several existing studies that use multimodal information to measure cognitive load (Son et al. 2013; Novak et al. 2011). We cannot compare the numerical results of our study with the numerical results of these studies because of differences in experimental design, populations, and measured signals. We can, however, compare our study with the existing studies to understand the effect of multimodal information in cognitive load measurement. Son et al. collected three modalities of information – physiological, gaze, and performance information, for cognitive load measurement in a driving simulator (Son et al. 2013). In their study, the best accuracy using the three-modality information was higher than the best accuracy using each single modality for cognitive load measurement. In an adaptive upper extremity rehabilitation task, Novak et al. showed that measuring cognitive load with physiological signals and task performance together can produce higher accuracy than using task performance or physiological signals, separately (Novak et al. 2011). In a mental arithmetic task, Hussain et al. found that multimodal fusion could increase the accuracy of cognitive load measurement when no affective interference was involved (M. S. Hussain et al. 2013). While these studies were not designed for individuals with ASD, our results are in line with the existing results in cognitive load measurement using multimodal information for TD individuals. To the best of our knowledge, no study fused multimodal information to measure cognitive load of individuals with ASD.

6.7.2. Decision level fusion and hybrid level fusion

We investigated the following research question in this work: which level of multimodal fusion can give the best accuracy in cognitive load measurement? In order to answer this question, we compared the best accuracies that can be achieved using different levels of fusion, including feature level fusion, decision level fusion, and hybrid level fusion, in cognitive load measurement. Table 44 summarizes the

best accuracies of the three multimodal fusion levels and shows that feature level fusion outperforms all other multimodal fusion levels in cognitive load measurement. Referring to previous literature in multimodal fusion, feature level fusion can achieve higher accuracies than decision level fusion due to the fact that feature level fusion utilizes the correlation among features from different modalities (Atrey et al. 2010). In our case, the effect of the correlation among features from different modalities can be seen from the best accuracy of hybrid level fusion with three sub-decisions. The best accuracy of hybrid level fusion with the three sub-decisions was achieved when eye gaze and EEG features were combined for one sub-decision, shown in Table 42. The correlation between eye gaze and EEG signals are significant (Dement and Kleitman 1957). The instance of hybrid level fusion utilizing this correlation achieved a higher accuracy than those that did not use this correlation.

Table 44 Comparison Between Different Levels of Fusion

	Feature level fusion	Decision level fusion	Hybrid level fusion
Best accuracy	84.43%	81.48%	83.42%

6.8. Conclusions and Future Research

6.8.1. Conclusions

ASD is a highly prevalent neurodevelopmental disorder. A novel VR-based driving system was presented for ASD intervention that could present driving scenarios with variable task difficulties to facilitate individualized learning. The primary contribution of this work is to systematically present the cognitive load measurements of individuals with ASD based on their eye gaze, EEG, peripheral physiology and performance data collected when they used the VR-based driving system, and to provide multimodal fusion schemes to more accurately measure cognitive load of these users. Feature level, decision level and hybrid level fusions demonstrate how multimodal information can be fused to measure cognitive load with increased accuracy. The model development for cognitive load measurement in this work is aimed at building a cognitively intelligent VR-based driving system. In the future, the difficulty

level of driving tasks will be adjusted in the cognitively intelligent VR-based driving system based on the research findings.

Our study has two distinct strengths that indicate the commercial viability of this system. First, it was tested with the intended target population, i.e., adolescents with ASD. Thus the system was acceptable and engaging to the target users. Second, the users used the driving simulator in a naturalistic way – they moved frequently and used it like a video game. As a result, the data was noisy and lost on occasions. Even then the cognitive load analysis was robust enough to predict cognitive load with a high accuracy in the presence of lost and noisy data. Thus we believe that this system will be commercially viable once cognitive load measurement mechanism presented in this work is integrated with the rest of the system.

6.8.2. Limitations and future research directions

There are several limitations of this research that need to be addressed in future work. First, we lost a relatively large quantity of data (20.56 % of all the data). The data were lost primarily due to participants' movements during driving, which was inevitable in the VR-based driving system aiming at training driving skills of adolescents with ASD in naturalistic conditions. One possible solution is detecting the valid data in real time in the cognitively intelligent VR-based driving system. If insufficient data for feedback is detected, the experiment could be extended in order to get more data.

Second, no multiple-class classification was analyzed in this work. We attempted binary classification as a starting point because it was simpler and in many cases, sufficient. However, in a more complex system, multi-class classification may yield richer results and should be investigated in the future.

Third, the data fusion method used in this work was limited. We combined sub-decisions for the final decision using a weighted average method. It is possible to use other methods to combine the sub-decisions, such as majority voting and classification algorithms (Atrey et al. 2010). We plan to explore different methods to combine sub-decisions in the future.

Finally, predefined, rather than randomized, difficulty levels were used in our experiments. Presenting randomized difficulty levels would be a better strategy for ultimately deciphering and analyzing potential

confounds associated with task difficulty level. We chose to present increasing difficulty levels in this initial pilot study in order to match a participant's expected skill increase with the higher levels of task difficulty. We will implement randomized difficulty levels in the future.

Even with above-mentioned limitations, we believe that this current work presents significant contributions towards developing cognitively intelligent VR-based driving systems that are robust, accurate, and useful for real-world applications indicating commercial viability in the near future. This system was explicitly designed for an ASD population who evidence both challenges with this functional adaptive skill (e.g., driving) and also historically found to evidence systemic, but complex heterogeneous impairments regarding information processing (e.g., working memory and executive functioning challenges, difficulties with social processing). We hypothesize that a multimodal fusion methodology capable of use within/across readily controllable intervention platforms (such as VR) could yield a tool for dramatically improving current modes of treatment. In this capacity the current findings will be used in future work developing a cognitively intelligent VR-based driving system. Its efficiency will be evaluated by comparing with a system without cognitive load-based feedback.

The generalizability of the training using our VR-based driving system will be evaluated in the real world in the future. A driving simulator, such as our VR-based driving system, is obviously not perfect for the on-road setting (Godley et al. 2002). However, it should be noted that driving behaviors of people in such kinds of simulators are similar to their driving behaviors in real-world driving (Keith et al. 2005). The speed patterns of people driving in a driving simulator were found to be similar to the speed patterns when driving in real world (Bella 2008). Traffic risk pattern, in terms of crash history, has also been shown to generalize from the simulator to the real world (Yan et al. 2008). The extant literature supports the usefulness of driving simulators. Evaluating the usefulness of training using our VR-based driving system for real world driving, in terms of speed-maintenance and error-reduction, will be carried out in future work.

6.9. References

Aditya, S., & Tibarewala, D. (2012). Comparing ANN, LDA, QDA, KNN and SVM algorithms in classifying relaxed and stressful mental state from two-channel prefrontal EEG data. *International Journal of Artificial Intelligence and Soft Computing, 3*(2), 143-164.

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review, 22*(4), 425-438.

Atrey, P. K., Hossain, M. A., El Saddik, A., & Kankanhalli, M. S. (2010). Multimodal fusion for multimedia analysis: a survey. *Multimedia systems, 16*(6), 345-379.

Bella, F. (2008). Driving simulator for speed research on two-lane rural roads. *Accident Analysis & Prevention, 40*(3), 1078-1087.

Bennetto, L., Pennington, B. F., & Rogers, S. J. (1996). Intact and impaired memory functions in autism. *Child development, 67*(4), 1816-1835.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research, 13*(1), 281-305.

Bishop, C. M. (2006). *Pattern recognition and machine learning* (Vol. 4, Vol. 4): springer New York.

De Clercq, W., Vergult, A., Vanrumste, B., Van Paesschen, W., & Van Huffel, S. (2006). Canonical Correlation Analysis Applied to Remove Muscle Artifacts From the Electroencephalogram. *Biomedical Engineering, IEEE Transactions on, 53*(12), 2583-2587, doi:10.1109/tbme.2006.879459.

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science, 38*(2), 105-134.

Dement, W., & Kleitman, N. (1957). Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and clinical neurophysiology, 9*(4), 673-690.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics, 7*(2), 179-188.

Gevins, A., & Smith, M. E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral cortex, 10*(9), 829-839.

Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., et al. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human factors: the journal of the human factors and ergonomics society, 40*(1), 79-91.

Godley, S. T., Triggs, T. J., & Fildes, B. N. (2002). Driving simulator validation for speed research. *Accident Analysis & Prevention, 34*(5), 589-600.

Hagan, M. T., Demuth, H. B., Beale, M. H., & De Jesús, O. (1996). *Neural network design* (Vol. 20): PWS publishing company Boston.

Hussain, M. S., Calvo, R. A., & Chen, F. (2013). Automatic cognitive load detection from face, physiology, task performance and fusion during affective interference. *Interacting with computers*, iwt032.

Hussain, S., Chen, S., Calvo, R. A., & Chen, F. Classification of Cognitive Load from Task Performance & Multichannel Physiology during Affective Changes. In *MMCogEmS: Inferring Cognitive and Emotional States from Multimodal Measures, ICMI 2011 Workshop, Alicante, Spain, 2011*

Keith, K., Trentacoste, M., Depue, L., Granda, T., Huckaby, E., Ibarguen, B., et al. (2005). Roadway human factors and behavioral safety in europe.

Kim, J., & André, E. (2008). Emotion recognition based on physiological changes in music listening. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 30*(12), 2067-2083.

Klem, G. H., Lüders, H. O., Jasper, H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. *Electroencephalogr Clin Neurophysiol, 52*(suppl.), 3.

Koelstra, S., Mühl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2012). Deap: A database for emotion analysis; using physiological signals. *Affective Computing, IEEE Transactions on, 3*(1), 18-31.

Koenig, A., Novak, D., Omlin, X., Pulfer, M., Perreault, E., Zimmerli, L., et al. (2011). Real-time closed-loop control of cognitive load in neurological patients during robot-assisted gait training. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on, 19*(4), 453-464.

Lagun, D., Manzanares, C., Zola, S. M., Buffalo, E. A., & Agichtein, E. (2011). Detecting cognitive impairment by eye movement analysis using automatic classification algorithms. *Journal of neuroscience methods, 201*(1), 196-203.

Lahiri, U., Warren, Z., & Sarkar, N. (2011). Design of a gaze-sensitive virtual social interactive system for children with autism. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on, 19*(4), 443-452.

Lin, Y., Tang, P., Zhang, W., & Yu, Q. (2005). Artificial neural network modelling of driver handling behaviour in a driver-vehicle-environment system. *International Journal of Vehicle Design, 37*(1), 24-45.

Liu, C., Agrawal, P., Sarkar, N., & Chen, S. (2009). Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *International Journal of Human-Computer Interaction, 25*(6), 506-529.

Liu, E. S., & Theodoropoulos, G. K. (2014). Interest management for distributed virtual environments: A survey. *ACM Computing Surveys (CSUR), 46*(4), 51.

Lushchekina, E., Podreznaya, E., Lushchekin, V., Novototskii-Vlasov, V. Y., & Strelets, V. (2013). Characteristics of the spectral power of EEG rhythms in children with early childhood autism and their association with the development of different symptoms of schizophrenia. *Neuroscience and Behavioral Physiology, 43*(1), 40-45.

Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record: Journal of the Transportation Research Board, 2138*(1), 6-12.

Meshkati, N., Hancock, P. A., Rahimi, M., & Dawes, S. M. (1995). Techniques in mental workload assessment.

Miller, S. (2001). Workload measures. *National Advanced Driving Simulator. Iowa City, United States*.

Mukaka, M. (2012). A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal, 24*(3), 69-71.

Narsky, I., & Porter, F. C. (2013). *Statistical analysis techniques in particle physics: Fits, density estimation and supervised learning*: John Wiley & Sons.

Novak, D., Mihelj, M., & Munih, M. (2012). A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing. *Interacting with computers, 24*(3), 154-172.

Novak, D., Mihelj, M., Ziherl, J., Olensek, A., & Munih, M. (2011). Psychophysiological measurements in a biocooperative feedback loop for upper extremity rehabilitation. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on, 19*(4), 400-410.

Olsen, A. (2012). The Tobii I-VT fixation filter. *Copyright© Tobii Technology AB*.

Ozonoff, S., & Strayer, D. L. (2001). Further evidence of intact working memory in autism. *Journal of autism and developmental disorders, 31*(3), 257-263.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist, 38*(1), 63-71.

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of educational psychology, 84*(4), 429.

Paas, F. G., & Van Merriënboer, J. J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351-371.

Pomplun, M., & Sunkara, S. Pupil dilation as an indicator of cognitive workload in human-computer interaction. In *Proceedings of the International Conference on HCI, 2003*

Rajendran, G., & Mitchell, P. (2007). Cognitive theories of autism. *Developmental review, 27*(2), 224-260.

Reimer, B., Fried, R., Mehler, B., Joshi, G., Bolfek, A., Godfrey, K. M., et al. (2013). Brief report: Examining driving behavior in young adults with high functioning autism spectrum disorders: A pilot study using a driving simulation paradigm. *Journal of autism and developmental disorders, 43*(9), 2211-2217.

Remington, A., Swettenham, J., Campbell, R., & Coleman, M. (2009). Selective attention and perceptual load in autism spectrum disorder. *Psychological Science, 20*(11), 1388-1393.

Safavian, S. R., & Landgrebe, D. (1990). A survey of decision tree classifier methodology.

Sarkar, N. Psychophysiological control architecture for human-robot coordination-concepts and initial experiments. In *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on, 2002* (Vol. 4, pp. 3719-3724): IEEE

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*(4), 469-508.

Schoor, C., Bannert, M., & Brünken, R. (2012). Role of dual task design when measuring cognitive load during multimedia learning. *Educational Technology Research and Development, 60*(5), 753-768.

Son, J., Oh, H., & Park, M. (2013). Identification of driver cognitive workload using support vector machines with driving performance, physiology and eye movement in a driving simulator. *International Journal of Precision Engineering and Manufacturing, 14*(8), 1321-1327.

Son, J., & Park, M. Estimating cognitive load complexity using performance and physiological data in a driving simulator. In *Adjunct Proceedings of the Automotive User Interfaces and Interactive Vehicular Applications Conference, 2011*

Steichen, B., Conati, C., & Carenini, G. (2014). Inferring Visualization Task Properties, User Performance, and User Cognitive Abilities from Eye Gaze Data. *ACM Transactions on Interactive Intelligent Systems (TiiS), 4*(2), 11.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123-138.

Wade, J., Zhang, L., Bian, D., Fan, J., Swanson, A., Weitlauf, A., et al. (2016 (In press)). A gaze-contingent adaptive virtual reality driving environment for intervention in individuals with autism spectrum disorders. *ACM Transactions on Interactive Intelligent Systems*.

Werling, D. M., & Geschwind, D. H. (2013). Sex differences in autism spectrum disorders. *Current opinion in neurology, 26*(2), 146.

Wu, D., Courtney, C. G., Lance, B. J., Narayanan, S. S., Dawson, M. E., Oie, K. S., et al. (2010). Optimal arousal identification and classification for affective computing using physiological signals: virtual reality Stroop task. *Affective Computing, IEEE Transactions on, 1*(2), 109-118.

Yan, X., Abdel-Aty, M., Radwan, E., Wang, X., & Chilakapati, P. (2008). Validating a driving simulator

using surrogate safety measures. *Accident Analysis & Prevention, 40*(1), 274-288.

Yannakakis, G. N., & Togelius, J. (2011). Experience-driven procedural content generation. *Affective Computing, IEEE Transactions on, 2*(3), 147-161.

CHAPTER VII. CONTRIBUTIONS AND FUTUER WORK

7.1. Contributions

7.1.1. Main Contributions

This dissertation describes my research on the design, development and evaluation of multi-user and intelligent Human-Computer-Interaction (HCI) systems for Autism Spectrum Disorder (ASD) intervention. Currently, cost and resource limitations impede access to effective ASD intervention. Previous work has shown that HCI systems hold promise as alternative ways of providing innovative, low-cost, and accessible clinical treatments for children with ASD. However, these investigations were based on rigid and limited interactions between users and computer programs. Such interactions demonstrate weak transfer of the trained skills to real world settings, which is the ultimate goal of therapies. Therefore, HCI intervention systems that could facilitate real-world interactions between multiple users are highly needed. In addition, the literature lacks efficient measurement strategies to index interactions within HCI systems. Consequently, manual coding of the interactions is necessary to understand users' behaviors for meaningful measurements. However, the human-coding methodology not only requires considerable human efforts but also limits the precision and scale-up of these paradigms (M Schmidt et al. 2011). As a result, there are pressing needs for both effective treatments to impact the neurodevelopmental trajectories of children with ASD and less burdensome measures to evaluate impacts of the treatments.

This research addresses these critical existing gaps in the literature. The main technical contributions of this work include, i) designing, developing, and applying Collaborative Virtual Environment (CVE) systems, which are computer-based, distributed, virtual spaces for multi-user interactions (Benford et al. 2001), to facilitate realistic interactions between real-users, and ii) exploring artificial intelligence methodologies to automatically measure users' behaviors in HCI-based intervention systems. In addition, this work also contributes to the science of ASD intervention by providing controllable and intelligent environments where different treatment paradigms can be accessed by multiple users in a flexible manner.

7.1.2. Technical Contributions

System design contribution

The first technical contribution of this work is the design and development of CVEs to encourage collaborations between real users. CVEs preserve the advantages of traditional HCI-based intervention systems but also facilitate realistic interactions between real users, increasing possible generalizability of learned skills to real-world settings. We developed CVEs to facilitate collaboration skills of children with ASD using collaborative puzzle games with multiple collaborative strategies, which required the children to talk with each other and move pieces together. Early studies in this area usually utilized a rule-based method to implement the collaborative strategies (Leman 2015; Benford et al. 2001). However, the rule based method could only model discrete actions. We applied a hybrid automaton to implement the collaborative strategies in our CVEs. Compared to the rule-based method, the hybrid automaton has the advantage to model both discrete variables and continuous actions of multiple users in the CVEs.

The feasibility of our CVE systems for children with ASD have been evaluated using several studies. In Chapter II, we developed a CVE, named CoMove, with a castle game and a set of seven tangram games, and equipped these games with collaborative strategies using the hybrid automaton. A study with seven TD/TD pairs and seven ASD/TD pairs demonstrated its feasibility with this population. In Chapter III, we transferred these games into a CVE on the Android platform and determined its feasibility using five ASD/TD pairs of children. Results of these studies indicated that CVEs can be used to encourage collaborations between children with ASD and their TD peers. In addition, these studies demonstrated the usability of the hybrid automaton in implementing collaborative strategies to encourage collaborations. As a result, this work contributes to the literature for the purpose of informing other researchers in designing HCI systems to encourage collaborations between real-users.

Contributions in measuring interactions in CVEs

The second set of technical contributions of this work relies on measuring peer-mediated interactions in CVEs. This type of measurement is essential to determining the CVE's impacts on users' social communication. The majority of existing CVEs for ASD intervention measure the users' interactions

based on their task-performance. However, understanding how the users verbally communicate and converse with one another is essential to the measurements. Existing work has relied upon a labor intensive human-coding methodology to understand the verbal communications (Matthew Schmidt et al. 2012). However, systems using this time-consuming methodology could not provide feedback in real-time. To address these limitations, we designed a measurement system that applied an intelligent agent to automatically index important aspects of the peer-mediated interactions in a CVE.

First, we designed an intelligent agent to address fundamental challenges in measuring peer-mediated interactions in CVEs, which are dynamic in nature and consist of open-ended verbal communications. In order to address these challenges, we developed an intelligent agent that could not only communicate and play games with users in a CVE, but also generate task-performance and verbal-communication features to represent the users' behaviors within the CVE. This intelligent agent was developed using a novel hybrid method, which combined a dialogue act classification and a finite state machine. The dialogue act classification classified users' natural language into one of several predefined dialogue acts, which are believed to be informative in indexing verbal-communication in CVEs (Caballé et al. 2011; Van Boxtel et al. 2000). The finite state machine combined users' verbal-communication and task-performance to generate speech responses and take game actions. This hybrid method allowed the intelligent agent to consistently interact with all users, as well as automatically generate meaningful features to measure these users' behaviors. As described in Chapter IV, a test study involving five children with ASD demonstrated the feasibility of the intelligent agent to communicate and collaborate with participants in a CVE.

Second, we proposed a framework to automatically measure users' communication skills and collaboration skills in a CVE to fill gaps in this area. The literature lacks efficient methods to measure users' behaviors in CVEs. In order to fill this gap, we proposed a framework to automatically measure these behaviors. This framework works in three steps. First, both task-performance and verbal-communication features were automatically generated using the intelligent agent. Second, the reliability of these features were evaluated using statistical analysis tests. Third, all the features were combined together to measure users' communication and collaboration skills with machine learning techniques.

We evaluated the feasibility of the framework using a user study with 20 pairs of children with ASD and TD children. Results of the study demonstrated the framework's ability to measure participants' both communication and collaboration skills in a CVE. This is the first attempt to automatically measure peer-mediated interactions in CVEs. This framework, if utilized, could reduce the time, costs, and efforts involved in measurements compared to traditional human-coding methodologies. In addition, it has the potential to enable real-time feedback for each individual to improve their learning efficiency. Although the framework was evaluated in a CVE with collaborative puzzle games, it could be transferred for measurements in other CVEs.

Contributions in cognitive load measurements

We also provided a framework to measure cognitive load of children with ASD using data fusion technologies. Cognitive load is believed to be a crucial factor for children with ASD to acquire knowledge and skills (Paas et al. 2003). Previous literature has analyzed eye gaze (Pomplun and Sunkara 2003), peripheral physiology (Liu et al. 2009), and electroencephalography data (Fan et al. 2015), respectively, for cognitive load measurements in the ASD population. We contributed to this area by combining these multimodal data to increase the measurement accuracy. Three data fusion strategies, i.e. feature-level fusion, decision-level fusion, and hybrid-level fusion, were explored in this study. Results indicated that multimodal fusion methods could outperform single modality classification in measuring cognitive load of individuals with ASD. In addition, we developed a novel method to find the optimized weights, which are the parameters used to fuse different modalities' results in the data fusion strategies. Compared to the traditional exhaustive search method (Koelstra et al. 2012), our method could significantly reduce computational load. In conclusion, this work contributes to the larger area of applying intelligent HCI systems to measure cognitive load of children with ASD.

7.1.3. Contributions to the Science of ASD Intervention

In addition to its technical contributions, this work also contributes towards the science of ASD intervention by providing controllable and intelligent environments wherein different intervention

paradigms can be assessed by multiple users in a flexible manner. Traditional paradigms often require significant time and effort-intensive burdens for implementation, suffer from limited availability in community settings, and ultimately demonstrate weak transfer of skills to real world settings (Weitlauf et al. 2014; Veenstra-VanderWeele and Warren 2015). The CVE-based intervention systems in this work offer a flexible alternative to conventional modalities of both in-vivo (e.g., social skill groups, peer-mediated programs) and technological intervention (e.g., confederate controlled HCIs, computerized skill programs) where multiple individuals can share and interact in a virtual space using network communication. Such technologically sophisticated systems are highly controllable, and can be adapted and structured in ways that mimic aspects of real-world interactions. As a result, the systems could tangibly impact the very nature of the collaborative interaction itself.

7.2. Future Work

Although this preliminary work is promising, future studies should address several important limitations. First, the sample size was relatively small, and the intervention duration for each work was relatively short. Therefore, the clinical impact of the proposed systems on everyday functioning of children with ASD is still unclear. Although the CVE-based intervention systems developed in this work demonstrated the potential to encourage collaborations within the systems, future studies should evaluate how the within-system interactions correlate with and potentially impact participants' skills in real life using a longitudinal clinical study with a large number of participants.

Second, current CVE-based intervention systems in Chapter II and Chapter III have the potential to facilitate verbal communication between real-users. These intervention systems without face-to-face communication have the advantages to simplify data analysis with emphasis on the verbal communication in preliminary studies. However, face-to-face communication is essential for real world interactions, improving which is the ultimate goal of treatment. Future systems should enable and assess aspects of face-to-face communication using a video chat functionality.

Third, the measurement system presented in Chapter V could only measure verbal-communication and task performance with 12 features. It did not assess other aspects of human behaviors that relate to peer-mediated interaction, such as eye gaze, body language, and facial expression. Additional work is needed to investigate how these features can be captured with separate computer programs, such as eye gaze recognition, gesture recognition, and emotion recognition, in order to understand the non-verbal communications.

Finally, we intend to utilize the real-time measurements in this work for a future adaptive system. The current intervention systems utilized performance-based feedback. We intend to incorporate verbal-communication-based feedback in order to foster communication skills based on these real-time measurements for children with ASD. In particular, the measurement results will be used to adapt system tasks for each individuals. Such an adaptive system will provide appropriate tasks for each child with ASD to practice his/her social communication skills in an efficient way based upon measured vulnerabilities.

7.3. References

Benford, S., Greenhalgh, C., Rodden, T., & Pycock, J. (2001). Collaborative virtual environments. *Communications of the ACM, 44*(7), 79-85.

Caballé, S., Daradoumis, T., Xhafa, F., & Juan, A. (2011). Providing effective feedback, monitoring and evaluation to on-line collaborative learning discussions. *Computers in Human Behavior, 27*(4), 1372-1381.

Fan, J., Wade, J. W., Bian, D., Key, A. P., Warren, Z. E., Mion, L. C., et al. A step towards EEG-based Brain computer interface for autism intervention. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015* (pp. 3767-3770): IEEE

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2012). Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing, 3*(1), 18-31.

Leman, P. J. (2015). How do groups work? Age differences in performance and the social outcomes of

peer collaboration. *Cognitive science, 39*(4), 804-820.

Liu, C., Agrawal, P., Sarkar, N., & Chen, S. (2009). Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *International Journal of Human-Computer Interaction, 25*(6), 506-529.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist, 38*(1), 63-71.

Pomplun, M., & Sunkara, S. Pupil dilation as an indicator of cognitive workload in human-computer interaction. In *Proceedings of the International Conference on HCI, 2003*

Schmidt, M., Laffey, J., & Stichter, J. Virtual social competence instruction for individuals with autism spectrum disorders: Beyond the single-user experience. In *Proceedings of CSCL, 2011* (pp. 816-820)

Schmidt, M., Laffey, J. M., Schmidt, C. T., Wang, X., & Stichter, J. (2012). Developing methods for understanding social behavior in a 3D virtual learning environment. *Computers in Human Behavior, 28*(2), 405-413.

Van Boxtel, C., Van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and instruction, 10*(4), 311-330.

Veenstra-VanderWeele, J., & Warren, Z. (2015). Intervention in the context of development: pathways toward new treatments. *Neuropsychopharmacology, 40*(1), 225.

Weitlauf, A. S., McPheeters, M. L., Peters, B., Sathe, N., Travis, R., Aiello, R., et al. (2014). Therapies for children with autism spectrum disorder.

APPENDIX

A. Results of Chapter V

A.1. Data analysis for ASD

Based on the framework discussed in Section 5.4, we also analyzed data of children with ASD. All the results of children with ASD are shown in Table 45, Table 46, and Table 47.

Table 45 Correlation between each system-generated feature and human ratings on a continuous scale

System-generated feature	Correlation between a feature and communication skills in HHIs	Correlation between a feature and collaboration skills in HHIs	Correlation between a feature and communication skills in HAIs	Correlation between a feature and collaboration skills in HAIs
Word frequency	0.6766**	0.2401*	0.3283**	0.1199
Request_color frequency	0.1858	0.0709	0.0762	0.0022
Provide frequency	0.5054**	0.2218	0.4521**	0.3460**
Direct_movement frequency	0.6118**	0.2250	0.2177	0.0298
Acknowledge frequency	0.3371**	0.1788	-0.0001	0.0928
Request_object frequency	0.1568	0.0807	-0.03724	-0.1880
Sentence frequency	0.7624**	0.3408**	0.4674**	0.3823**
Success frequency	-0.1805	0.4276**	0.3784**	0.5378**
Failure frequency	-0.0892	-0.4252**	-0.4896**	-0.5977**
Collaboration time	0.0224	0.4208**	0.3041	0.4289**
Dragging time	0.0493	0.0610	-0.3727**	-0.3281**
Collaborative movement ratio	-0.2298	0.2850	0.1018	0.1668

Note: ** indicates a p value less that .001; * indicates a p value less than .05

Table 46 Correlation between a system-generated feature and human ratings on a binary scale

System-generated feature	Correlation a feature and communication skills in HHIs	Correlation a feature and collaboration skills in HHIs	Correlation a feature and communication skills in HAIs	Correlation a feature and collaboration skills in HAIs
Word frequency	0.6085**	0.2189*	0.2184	-0.0153
Request_color frequency	0.2097	0.0854	0.0744	0.0161
Provide frequency	0.3894	0.1224	0.3523**	0.2465*
Direct_movement frequency	0.5703**	0.2248*	0.1358	-0.0288
Acknowledge frequency	0.2547	0.1056	0.1338	0.1043
Request_object frequency	0.1460	0.0873	-0.03502	-0.1819
Sentence frequency	0.6550**	0.2625**	0.4638**	0.2573*
Success frequency	-0.1384	0.3065**	0.3236**	0.3882**
Failure frequency	-0.1160	-0.3317**	-0.4145**	-0.4802**
Collaboration time	0.0474	0.3052**	0.3265**	0.3128**
Dragging time	-0.0433	0.0229	-0.2865	-0.2537
Collaborative movement ratio	-0.1510	0.1949	0.1540	0.1429

Note: ** indicates a p value less that .001; * indicates a p value less than .05

Table 47 Accuracy of measuring communication skills and collaboration skills

Index	Which skills to measure?	Data sample size (High level /Low level)	Accuracy of balanced data	Accuracy of all data
1	Communication skills in HHIs	84/71	80.95%	0.8008
2	Collaboration skills in HHIs	119/36	73.61%	0.8256
3	Communication skills in HAIs	109/28	80.37%	0.7958
4	Collaboration skills in HAIs	100/37	77.03%	0.7740

A.2. Data analysis for TD

Based on the framework discussed in Section 5.4, we also analyzed data of TD children. Table 48, Table 49, and Table 50 shows results of TD children.

Table 48 correlation between a system-generated feature and human ratings on a continuous scale

System-generated feature	Correlation between a feature and communication skills in HHIs	Correlation between a feature and collaboration skills in HHIs	Correlation between a feature and communication skills in HAIs	Correlation between a feature and collaboration skills in HAIs
Word frequency	0.6389**	0.3117*	0.5905**	0.2562*
Request_color frequency	0.0967	0.1243	0.1473	-0.0281
Provide frequency	0.4504**	0.2335*	0.3426*	0.3107*
Direct_movement frequency	0.6050**	0.3596*	0.2378	-0.0322
Acknowledge frequency	0.3935*	0.1809	-0.2022	-0.0699
Request_object frequency	-0.0077	-0.1338	-0.0986	-0.0577
Sentence frequency	0.7623**	0.4467**	0.2794**	0.1838
Success frequency	-0.1861	0.4219**	0.0334	0.1804
Failure frequency	-0.0733	-0.4309**	-0.2368*	-0.2280**
Collaboration time	-0.0299	0.4306**	-0.0438	0.0569
Dragging time	0.1539	0.1507	0.0205	0.0753
Collaborative movement ratio	-0.2744	0.1898	-0.1684	-0.0309

Note: ** indicates a p value less that .001; * indicates a p value less than .05

Table 49 correlation between a system-generated feature and human ratings on a binary scale

System-generated feature	Correlation between a feature and communication skills in HHIs	Correlation between a feature and collaboration skills in HHIs	Correlation between a feature and communication skills in HAIs	Correlation between a feature and collaboration skills in HAIs
Word frequency	0.5256**	0.2487*	0.2253*	0.0369
Request_color frequency	0.0650	0.0853	0.0504	0.0292
Provide frequency	0.3747**	0.2022*	0.2275*	0.1663
Direct_movement frequency	0.4686**	0.2847	0.0171	-0.1748
Acknowledge frequency	0.3787**	0.1691	0.0435	0.0231
Request_object frequency	0.0209	-0.1631	0.0146	0.0164
Sentence frequency	0.6336**	0.3578**	0.2782	0.0642
Success frequency	-0.1613	0.3237**	0.0792	0.0601
Failure frequency	0.0305	-0.3870**	-0.0814	-0.0673
Collaboration time	-0.0796	0.3768**	-0.1678	-0.0249
Dragging time	0.1624	0.0908	0.0788	0.0502
Collaborative movement ratio	-0.2347*	0.1083	-0.2768*	-0.0336

Note: ** indicates a p value less that .001; * indicates a p value less than .05

Table 50 the accuracy of measuring communication skills and collaboration skills in HHIs

Index	Which skills to measure?	Data sample size (good/not good)	Accuracy of balanced data	Accuracy of all data
1	Communication skills in HHIs	110/45	0.8117	.7999
2	Collaboration skills in HHIs	117/38	0.7628	0.7428
3	Communication skills in HAIs	4/136	--*	0.9712
4	Collaboration skills in HAIs	5/135	--*	0.8256

Note: * indicate that the data sample size is too small to build a SVM-RBF model

A.3. Data analysis for different games

We combined all the data of Game_1, Game_2, and Game_3 to be group1, all the data of Game_4, Game_5, and Game_6 as group2; and all the data of Game_7, Game_8, and Game_9 as group3. The characteristics of these games are shown in Table 20. Then, we calculated the correlations between each system-generated features and communication skills and communication skills, respectively, for each group. Table 51, Table 47, Table 48, and Table 54 shows all the correlations of these groups.

Table 51 correlations between features and continuous communication skills of three games

System-generated feature	Correlation between each feature and continuous communication skills of group1 in HHIs	Correlation between each feature and continuous communication skills of group2 in HHIs	Correlation between each feature and continuous communication skills of group3 in HHIs
Word frequency	0.7804**	0.7993**	0.7596**
Request_color frequency	0.1462	0.3345**	0.2176*
Provide frequency	0.5527**	0.6138**	0.5138**
Direct_movement frequency	0.7127**	0.7220**	0.7072**
Acknowledge frequency	0.3708**	0.3972**	0.1569
Request_object frequency	0.0793	0.0697	0.1153
Sentence frequency	0.7815**	0.8334**	0.7337**
Success frequency	0.0303	-0.2907	-0.0822
Failure frequency	-0.1907	0.1742	-0.1537
Collaboration time	0.0953	-0.1860	-0.0476
Dragging time	0.3211*	0.2898*	0.2089*
Collaborative movement ratio	0.1805	-0.0931	-0.0206

Table 52 correlations between features and continuous collaboration skills in different games

System-generated feature	Correlation with each feature and continuous collaboration of group1 in HHIs	Correlation with each feature and continuous collaboration of group 2 in HHIs	Correlation with each feature and continuous collaboration of group 3 in HHIs
Word frequency	0.4453**	0.2874*	0.1984
Request_color frequency	-0.0130	-0.0327	0.2586
Provide frequency	0.3153*	0.1710	0.0994
Direct_movement frequency	0.4330**	0.2240	0.1800
Acknowledge frequency	0.2952	0.1599	0.0049
Request_object frequency	-0.0843	-0.0258	-0.0692
Sentence frequency	0.5533**	0.3583**	0.2855*
Success frequency	0.5487**	0.4593**	0.5960**
Failure frequency	-0.4482**	-0.3421**	-0.3935**
Collaboration time	0.4399**	0.3379**	0.4217**
Dragging time	0.1944	0.1334	0.2283
Collaborative movement ratio	0.4356**	0.1490	0.3329*

Table 53 correlations between features and continuous communication skills of three games

System-generated feature	Correlation between each feature and continuous communication skills of group1 in HAIs	Correlation between each feature and continuous communication skills of group2 in HAIs	Correlation between each feature and continuous communication skills of group3 in HAIs
Word frequency	0.5611**	0.6027**	0.4752**
Request_color frequency	0.0693	0.0412	0.2442*
Provide frequency	0.3203*	0.4826**	0.5784**
Direct_movement frequency	0.1440	0.2013*	0.1489
Acknowledge frequency	-0.1643	-0.0262	-0.0558
Request_object frequency	-0.1970	-0.2876	-0.2982
Sentence frequency	0.1968	0.3563**	0.4081**
Success frequency	0.2429*	0.2306*	0.3598**
Failure frequency	-0.3211*	-0.4063**	-0.4296**
Collaboration time	0.2874*	0.2008*	0.2507*
Dragging time	0.2215	0.0280	0.0681
Collaborative movement ratio	0.2058	0.1408	0.1706

Table 54 correlations between features and continuous collaboration skills in different games

System-generated feature	Correlation with each feature and continuous collaboration of group1 in HAIs	Correlation with each feature and continuous collaboration of group 2 in HAIs	Correlation with each feature and continuous collaboration of group 3 in HAIs
Word frequency	0.3296*	0.3757*	0.2101*
Request_color frequency	0.0409	-0.1142	-0.0076
Provide frequency	0.3552*	0.4092**	0.4875**
Direct_movement frequency	0.1184	0.1464	0.1095
Acknowledge frequency	-0.2574	-0.0488	-0.0754
Request_object frequency	-0.2188	-0.2487	-0.2761
Sentence frequency	0.2737*	0.2920*	0.4084**
Success frequency	0.3949**	0.3901**	0.5418**
Failure frequency	-0.4794**	-0.4169**	-0.4839**
Collaboration time	0.4750**	0.3343**	0.4695**
Dragging time	0.4056**	0.0272	0.1030
Collaborative movement ratio	0.3408*	0.3160*	0.3868**