Design of a Virtual Reality-based Collaborative Activities Simulator (ViRCAS) to Support Teamwork in Workplace Settings for Autistic Adults

Ashwaq Z. Amat, Deeksha Adiani, Mahrukh Tauseef, Michael Breen, Spencer Hunt, Amy Swanson, Amy S. Weitlauf and Nilanjan Sarkar, *Senior Member, IEEE*

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

***Index Terms*— Autism, Automatic speech recognition, Collaborative tool, Gaze tracking, Human computer interaction, Intelligent system, Teamwork, Virtual reality**

I. Introduction

Manuscript received xxx xx, 20xx; revised xxx xx, 20xx; accepted xxx xx, 20xx. This work was supported by NSF grants 1936970 and 2033413. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

A.Z. Amat and M. Tauseef are with Department of Electrical and Computer Engineering, Vanderbilt University, Nashville, TN, USA. (e-mail: ashwaq.zaini.amat.haji.anwar@vanderbilt.edu; mahrukh.tauseef@vanderbilt.edu).

D. Adiani is with Department of Computer Science, Vanderbilt University, Nashville, TN, USA. (e-mail: deeksha.m.adiani@vanderbilt.edu).

M. Breen and S. Hunt are with Department of Mechanical Engineering, Vanderbilt University, Nashville, TN, USA. (e-mail: michael.breen@vanderbilt.edu).

A. Swanson is with the Treatment and Research Institute for Autism Spectrum Disorders, Nashville, TN, USA. (e-mail: amy.r.swanson@vumc.org).

A. S. Weitlauf is with Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, USA. (e-mail: amy.s.weitlauf@vumc.org).

N. Sarkar is with Departments of Mechanical Engineering, Electrical and Computer Engineering, and Computer SCience, Vanderbilt University, Nashville, TN, USA. 37212. (e-mail: nilanjan.sarkar@vanderbilt.edu).

AUTISM spectrum disorder (ASD) [1] impacts social communication and interaction as well as patterns of behavior and sensory processing [2]. One in 44 children and 1 in 45 adults are diagnosed with ASD each year in the US [3], [4] with more than 70,000 autistic children reached adulthood each year [5]. Differences in social communication and interaction can impede autistic adults as they attempt to secure and retain employment [6]. The unemployment rate for individuals with ASD is between 50% and 85%, the highest compared to other types of disabilities [7] with an estimated total lifetime cost of care for an autistic individual between $2.4 million and $3.2 million [8], stemming from unemployment and adult care costs [9].

Autistic adults may have many workplace-relevant talents [10]–[12], such as attention to detail [12]. However, differences in communication and social interaction skills relative to colleagues without ASD ("neurotypical") can impact employment opportunities that require a high level of teamwork [13]. In general, teamwork skills are associated with improved productivity and workplace performance [14], are among the core skills sought by employers, and can influence hiring decision [15]. Companies such as Microsoft and Specialsterne have started using a non-traditional interview process for autistic candidates to assess teamwork using Lego Mindstorm group projects [16] and Minecraft [17]. Therefore, supporting autistic adults to acquire work-relevant teamwork skills may contribute to improved workplace social communication skills [18], problem-solving skills [19], and self-confidence [20], while at the same time increase the opportunities to obtain meaningful work that aligns with their strengths and interests.

One way to assess and support teamwork skills development is simulation-based training (SBT), which enables individuals to engage in a shared social, cognitive, and behavioral process pertaining to a collaborative task [21]. Although existing SBT programs have positively impacted teamwork skills development [20], [22], these programs can be tedious, resource-straining, and costly [23], thus driving the need for technology-based solution. Over the last decade, the use of human-

[1]We are using identity-first language in this paper due to the preference of autistic individuals and their families on this disability-related terminology [1]

computer interaction (HCI) technology has shown promise by providing lower-cost, engaging interactions that improve accessibility [24]. Virtual Reality (VR) has been used to simulate real-world scenarios at a lower cost [25]. VR-based systems have shown potential for teaching both autisitic and neurotypical individuals new social and technical skills [26]–[30]. However, conventional VR-based systems are limited to single-user and unable to support the complex back and forth human-human interactions important for teamwork skills training. Additionally, overreliance on virtual interactions may limit generalizability and success in real-world tasks [31].

Effective teamwork requires collaboration among individuals working together, making collaboration an important indicator of teamwork performance [32]. A collaborative virtual environment (CVE) extends the benefits of conventional VR technology by supporting multi-user interaction within the same shared virtual space, allowing users to naturally communicate with each other [33], potentially increasing generalizability of learned skills to real world. Several recent CVE-based interactions have been promising. For example, a collaborative motor skill training system for autistic children showed increased in both motor and social skills [33]. A virtual learning environment to foster social cognitive skills for autistic young adults resulted in significant improvements in emotion recognition [34]. Social MatchUP, an immersive CVE-based system for adults with neurodevelopment differences like ASD, reported a significant improvement [35]. However, there are no CVE-based studies to our knowledge on social interactions within the employment landscape, which can be different from everyday social interactions.

Complex social skills such as teamwork can be challenging to assess [36]. Existing methods of assessment still rely heavily on human observations [34], [37]. Fortunately, studies on collaborative learning and communications can be leveraged to objectively assess teamwork skills [38]. Furthermore, advancements in HCI and sensors technologies have paved the way for the use of multimodal data to provide a reliable assessment of human behavior [39], through quantitative measures of several dimensions of collaboration [40].

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

The current work substantially expands our previous conference paper [41] in terms of 1) expansion of system interactivity: we incorporated audio and visual communication channels within the CVE that allow the users to see and talk to each other; 2) introduction of a new collaborative task: Task 3 in Section II-A-3 and the addition of difficulty levels in all tasks; and 3) classifying and assessing collaboration using multimodal data from a human participant study: we present a new human-subject study with 12 pairs of participants (6 ASD-NT pairs and 6 NT-NT pairs). The remainder of the paper is organized as follows: Section II presents the system design and the system architecture. Section III describes the experimental setup followed by Section IV, which presents the results of the study. Finally, Section V discusses the results and addresses the potential and limitations of the current study.

II. System Design

A. Collaborative Tasks Design Principles

1) Stakeholder-driven Universal Design of Collaborative Tasks: We employed a participatory design process where we engaged with stakeholders and end-users from various backgrounds to design meaningful collaborative tasks: industry representatives from 2 companies, a certified behavioral interventionist, 2 career counselors from 2 vocational rehabilitation centers, and 3 autistic adults. Stakeholders were involved in both the design and development stages of the collaborative tasks. In the design stage, we conducted multiple discussion sessions with the stakeholders to select tasks that are collaborative and include interactions that are suitable in a workplace environment. For example, a puzzle game task can be collaborative, but might not involve workplace-related interactions. The collaborative tasks selection were driven by employment-related studies for autistic individuals: a) a PC Assembly task [42], b) a Fulfillment Center task [25], and c) a Furniture Assembly task [43]. These tasks elicited teamwork-relevant behaviors between two users, could be designed at varying difficulty levels, and involved workplace-related interactions. Additionally, we incorporated universal design principles into our collaborative tasks to create a system that can be used by individuals with different abilities [44].

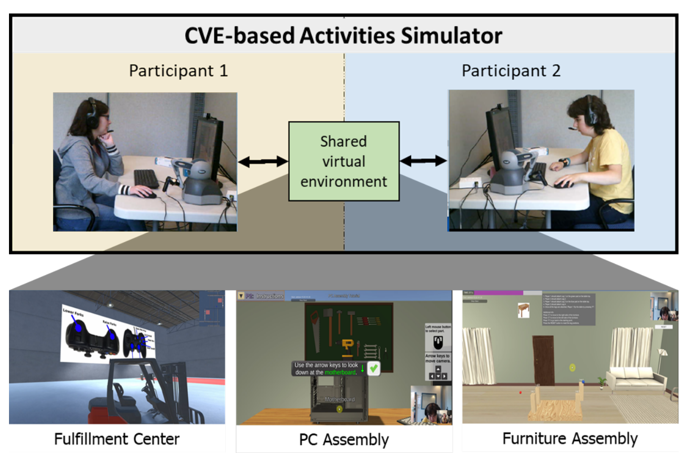

Fig. 1. ViRCAS setup and snapshots of the three collaborative tasks

In the development stage, we recruited 3 autistic adults and 3 neurotypical adults. Each ASD-NT pair tested the initial version of the collaborative tasks while being observed by two expert behavioral interventionists with prior experience

TABLE I
DIMENSIONS OF COLLABORATION

No.	Dimensions	Definition (The task should allow...)
1	Sustaining Mutual Understanding	Participants to share ideas and show mutual understanding.
2	Dialogue management (Turn-taking)	Participants to engage in back-and-forth communication and activities.
3	Information Pooling	Participants to share information with each other.
4	Reaching Consensus (Decision making)	Participants to discuss and agree with each other
5	Task Division	Participants to discuss and coordinate their actions within the task
6	Time Management	Participants to monitor and be aware of time restrictions in the task
7	Technical Coordination	Participants to handle technical dependencies of the task
8	Reciprocal Interaction	Participants to progress at the same pace
9	Individual Task Orientation	Participants to perform individual actions independently

in real-world teamwork tasks, who then commented on task suitability and made suggestions for improvement to aligh with real-world supports. We made several changes based on feedback from the interventionists and participants. First, a need was identified for a structured instruction. As a result, we developed a tutorial level that provided step-by-step instructions. Second, participants found that the virtual objects were difficult to manipulate. To address this concern, we simplified the object manipulation function. Third, at times, participants were not sure what they needed to do. We added visual cues that made it easier for participants to know where to go or which objects to move.

2) Collaboration dimensions for Collaborative Activities: Based on literature related to dyadic interactions and collaboration [38], [40], we defined and incorporated 9 dimensions of collaboration into the tasks. These dimensions use both verbal and non-verbal communications that can be quantitatively measured to represent the quality of collaboration between the participants. Table I lists the dimensions and their definitions.

3) Tasks Descriptions: First, we will describe the overview and setup of the collaborative activities simulator. Two participants in different physical locations accessed a shared virtual environment from their respective computers as illustrated in Figure 1. Each participant used the input device to interact with their virtual environment, a headphone with a microphone to communicate with their partner, and a webcam to see each other. After they were connected to the same virtual environment, participants could communicate with each other through an audio and video streaming component embedded within the virtual environment. They were asked to complete 3 levels of each task (see Table II for task and level descriptions).

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

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

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

B. CAS Architecture

1) Input Devices: One principle of universal design is perceptible information [44], which supports multiple methods of communication between users and the system. We employed three types of input devices with varying characteristics for each of the task as presented in Table VIII to explore their benefits. In the PC Assembly task the participants used a keyboard and mouse to move the virtual hardware. In the Fulfillment Center task the participants used a Logitech Gamepad [45]. Participants used the directional pad to drive the forklift in the virtual warehouse and the directional button to change the height of the fork when picking up a crate. In the Furniture Assembly task the participants used a haptic device [46] for greater immersion.

2) CVE Modules and Communication Network: Figure 2 illustrates the system interaction diagram and architecture. The ViRCAS was created using a virtual game development software, Unity [47].The Network Communication Module handles the connection of two participants to the same virtual environment. This module also manages real-time audio and video interaction. Virtual objects' synchronization was achieved using a Unity plugin called Mirror [48], while the audio and video data streaming were accomplished using WebRTC [49]. Task-related data are transmitted between the two computers in packets. Mirror uses Transmission Control Protocol (TCP) to send information between the two computers. TCP ensures data transmission from the source are correctly delivered to the target, in the right order, resulting in a synchronized shared environment. Although TCP assures data delivery to the target, the latency is slightly higher and could result in delay. However, for ViRCAS, the latency does

TABLE II
COLLABORATIVE TASKS LEVELS

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

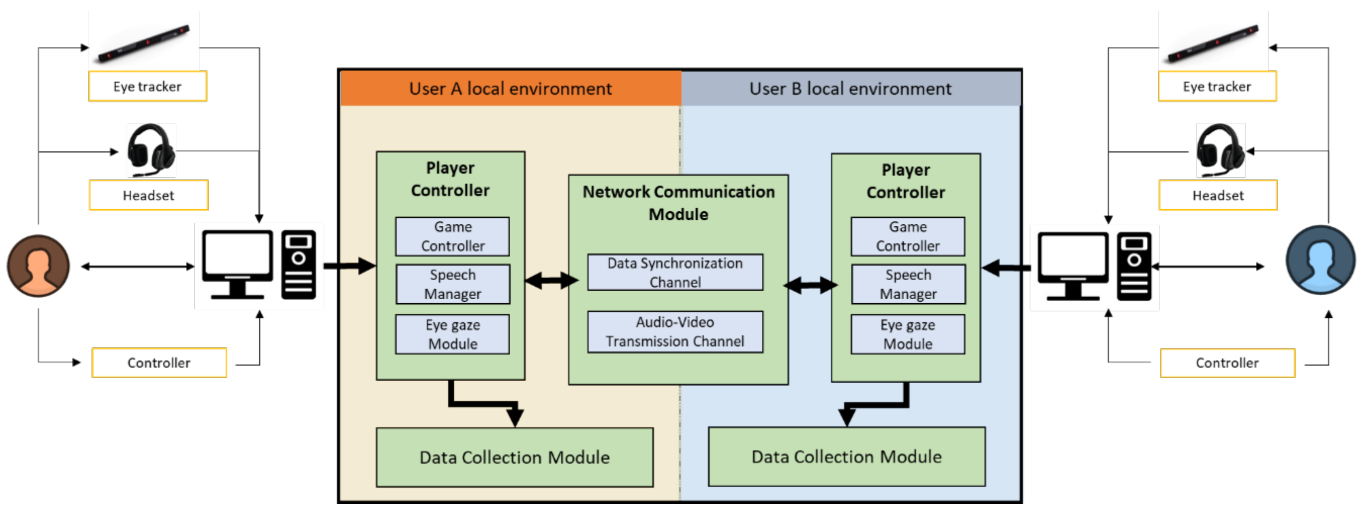

Fig. 2. Architecture of the collaborative system

TABLE III
INPUT DEVICES SPECIFICATION

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

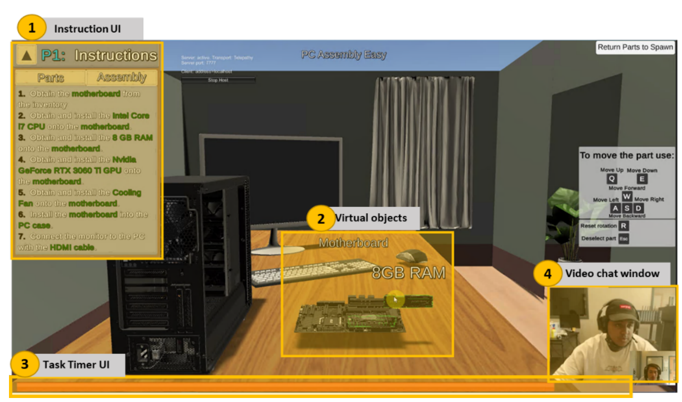

Fig. 3. Example of Region of Interests (ROIs) for PC Assembly Task

not significantly affect task interaction since our tasks do not require instantaneous updates. WebRTC uses User Datagram Protocol (UDP) for audio and video transmission prioritizes latency over data accuracy.

Next, the Player Controller component manages the use of multiple peripheral devices by participants to interact with the virtual environment. Task-related data collected in this module are sent over to the Network Communication Module. There are three sub-modules within the Player Controller component. First, the Game Controller manages input device manipulation of virtual objects. The Player Controller component manages the input devices and keeps track of the task time and task progression. Second, the Speech Manager processes participants' speech. The spoken words from both ASD and NT participants are transcribed into text in real-time using Microsoft Azure's Speech-to-Text service [50]. In Unity, we created a continuous listener function that captures any speech and sends it over to Azure API. Upon receiving the data, Azure proceeds to transcribe each word it received and grouped the words as one utterance. One utterance ends when silence was detected or a maximum of 15 seconds of audio was processed [51]. We can determine the number of words used in each utterance and the duration of the utterance with the transcribed speech. The final sub-module is the Eye Gaze Module which detects participants' eye gaze on the computer screen using a TobiiEyeX eye tracker [52]. Even though the sampling frequency of TobiiEyeX is comparatively low at 60 Hz, it is sufficient to detect gaze fixation on screen [53]. We utilized a Tobii Unity Eye Tracking SDK [54] to: 1) continuously capture gaze points, and 2) capture gaze fixations on pre-defined region of interests (ROIs) and virtual objects when a gaze duration of approximately 200 ms is detected. Figure 3 presents an example of ROIs for the PC Assembly task. Finally, the controller data, speech data, detected gaze points, and the ROIs were recorded together with the timestamps and sent to the Data Collection Model.

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

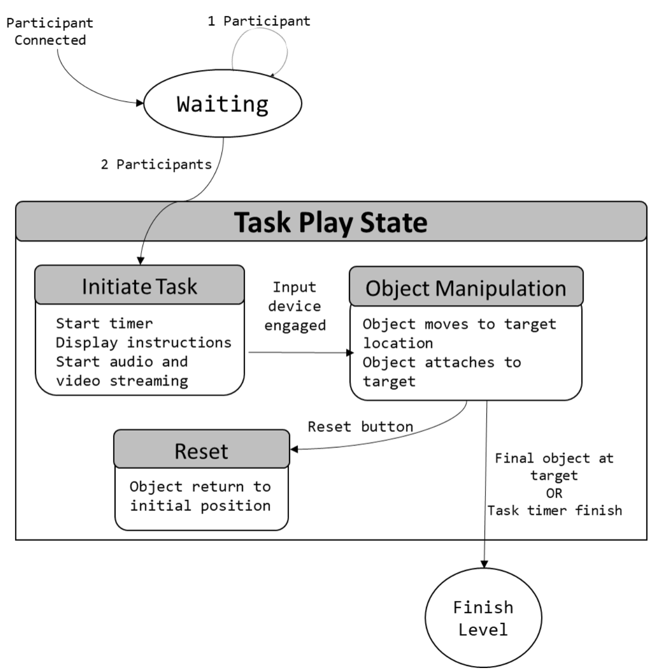

Fig. 4. Generic finite state machines for all tasks

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

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

III. Experimental Design

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

A. Participants and Protocol

We recruited 6 individuals with ASD and 18 NT individuals (ages: 16 – 30 years; mean age: 23.4 years) to participate in the study. Participants with ASD were recruited from a large research registry maintained by the Vanderbilt Kennedy Center of individuals previously diagnosed with ASD by licensed clinical psychologists. The NT participants were recruited from local community through regional advertisement. We then divided the participants into two groups: 6 ASD-NT pairs, and 6 NT-NT pairs.

Table IV shows the characteristics and current level of ASD symptoms of all participants as measured by the Social Responsiveness Scale, Second Edition (SRS-2) [55]. Note that SRS-2 T-scores of 66 and above reflect at least moderately elevated symptoms of ASD, while T-scores of 59 and below reflect little-to-no evidence of ASD.

Each pair of participants attended a one-time visit to the laboratory that lasted approximately 90 minutes. They were seated in two different experiment rooms and accessed ViRCAS from local area network (LAN) that ensured data

TABLE IV
CHARACTERISTICS OF PARTICIPANTS

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

security and privacy. Before the participants began their session, consents and assents from the participants' guardians and the participants themselves were obtained, respectively. Participants completed all levels (i.e., Tutorial, Easy, and Hard) of the PC Assembly, Furniture Assembly, and Fulfillment Center task.

IV. RESULTS

A. Acceptability of the Collaborative Tasks and Input Devices

We asked participants a total of 24 questions using a 10 point-Likert scale to get their feedback on each task, the input devices, and the overall acceptability of the system. Table V groups the questionnaire into three main categories and reports the mean and standard deviation of the responses for autistic and NT participants.Scores indicated that all tasks were acceptable to both autistic and NT participants.As for input device preference, participants with ASD preferred the gamepad the most while NT participants preferred keyboard and mouse usage. The haptic device was the least preferred device in both groups. However, participants did provide positive verbal responses on the force feedback of the haptic device during use.

B. Dialogue Acts Classification Results

We analyzed the transcribed speech data to better understand the context of the conversation. Table VI lists the annotation scheme adapted from a verbal behavior coding scheme used to classify speech [38]. Two annotators labeled the transcribed conversation between the participants using the coding scheme and reached an inter-annotator agreement of 95%. The annotators reconciled their differences to reach a final agreement of 100%. We used an unpaired t-test to evaluate any differences in the utterance pattern between the two groups.

As shown in Table VII, we found statistically significant differences in five types of dialogue acts between the ASD-NT group and NT-NT group. Pairs in the NT-NT group uttered more acknowledgements ('Acks', p-value 0.031), used more negative words ('Neg', p-value = 1.389e-5), asked more questions ('Ques', p-value = 0.0009), and instructive utterances ('Inform', p-value = 0.0001), while pairs in ASD-NT group used descriptive words more ('Desc', p-value =0.009) compared to NT-NT group. Figure 5 illustrates noticeable differences in the dialogue acts percentage between the groups.

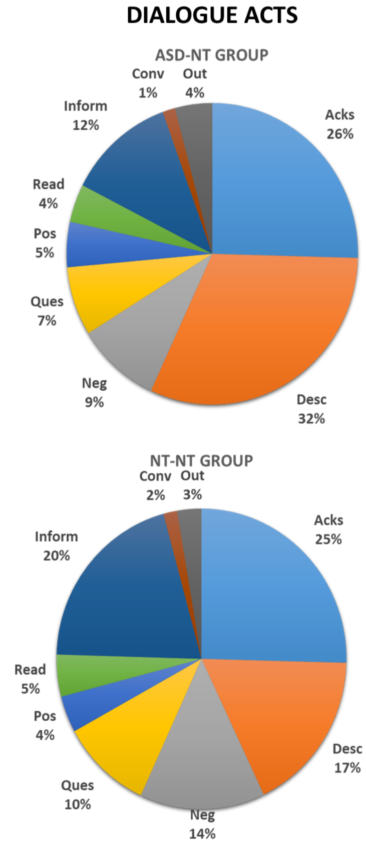

Fig. 5. Dialogue acts percentages in ASD-NT group and NT-NT group.

C. Utterances Analysis Results

We analyzed the utterances by grouping the number of utterances into Easy and Hard levels as we wanted to observe the impact of increasing the difficulty level on collaboration. We found that number of utterances per level increased as the difficulty level increased for all participants as shown in Figure 6.

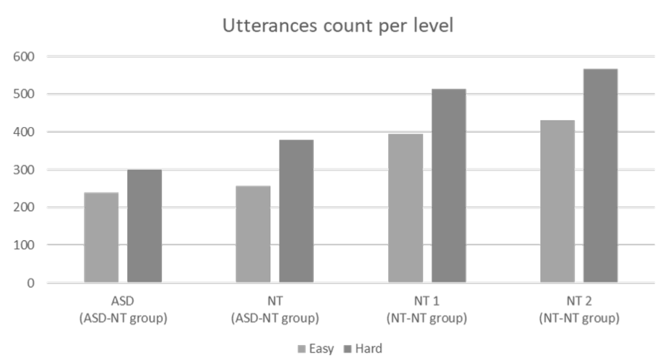

Fig. 6. Number of utterances increased for all participants as difficulty level increased

D. Gaze Duration Results

We analyzed the participants' gaze by calculating the duration of the gaze on the ROIs. Gaze duration that lasted

TABLE V
QUESTIONNAIRE SCORE

Questions	ASD (N = 6) Mean (SD)	NT (N = 18) Mean (SD)
Collaborative Tasks		
How confident did you feel throughout the task?	7.94 (2.50)	7.00 (2.72)
How comfortable did you feel overall with the task?	8.38 (1.78)	7.48 (2.22)
How comfortable did you feel interacting with your partner?	9.52 (1.23)	9.34 (1.16)
How comfortable did you feel when the task was challenging?	8.47 (1.88)	6.95 (2.43)
Input Devices		
How comfortable did you feel using the haptic device to move parts around?	6.66 (3.07)	4.88 (2.51)
How comfortable did you feel using the keyboard and mouse to move parts around?	7.16 (2.31)	8.11 (1.81)
How comfortable did you feel using the gamepad to move parts around?	8.83 (1.60)	7.44 (2.57)
ViRCAS System		
How much do you agree with the following: "Practicing with this system would help me work with others better."	9.50 (1.22)	8.11 (2.39)
How much do you agree with the following: "If it was available, I would use this system to practice my teamwork skills"	9.66 (0.81)	7.00 (3.04)

TABLE VI
DIALOGUE ACTS DEFINITIONS

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

TABLE VII
COMPARISON OF DIALOGUE ACTS CLASSIFICATION ACROSS GROUPS

Dialogue Act Labels	ASD-NT Mean (SD)	NT-NT Mean (SD)	t-test (p-value)
Acks	12.142 (82.304)	15.143 (58.924)	**1.889 (0.031)***
Desc	15.107 (153.188)	10.375 (67.002)	**-2.387 (0.009)***
Neg	4.214 (15.553)	8.393 (34.897)	**4.402 (1.389e-5)***
Pos	2.321 (6.986)	2.518 (6.509)	0.4001 (0.345)
Ques	3.625 (16.420)	6.107 (17.406)	**3.194 (0.0009)***
Read	1.857 (7.761)	2.589 (7.083)	1.422 (0.079)
Inform	5.518 (58.036)	12.286 (120.68)	**3.788 (0.0001)***
Conv	0.714 (0.826)	0.964 (1.089)	1.351 (0.09)
Out	1.964 (3.344)	1.582 (2.989)	-1.132 (0.13)

TABLE VIII
PARTICIPANTS' GAZE DURATION

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

approximately 250 ms was considered a "fixations gaze". Table 3 compares the average gaze fixations duration between autistic and NT participants. An unpaired t-test showed statistically significant differences in gaze fixations duration for the Fulfillment Center task; NT participants gazed 3 times longer at the virtual objects compared to participants with ASD (p-value = 0.012). Other tasks did not show any significant differences.

E. Observation of Dimension of Collaboration

To determine whether ViRCAS captured the dimensions of collaboration (Table I) from the multimodal data, we computed the occurrence of individual dimension from every 30s of sampled data. Figure 7 presents the dimensions of collaboration pattern in both groups. Both ASD-NT pairs and NT-NT pairs showed similar collaborative pattern for mutual understanding, dialogue management, information pooling, and consensus. However, pairs in the NT-NT group showed more reciprocal interaction and less technical coordination and individual motivation compared to the pairs in the ASD-NT group. In both groups, we did not observe sufficient task division and time management dimensions.

V. DISCUSSION

We designed and completed a feasibility study of ViRCAS, a novel collaborative activities simulator within CVE. The objectives of the study were to 1) assess the usability of ViRCAS among individuals with ASD, 2) assess the ability of the collaborative tasks to support teamwork, 3) observe

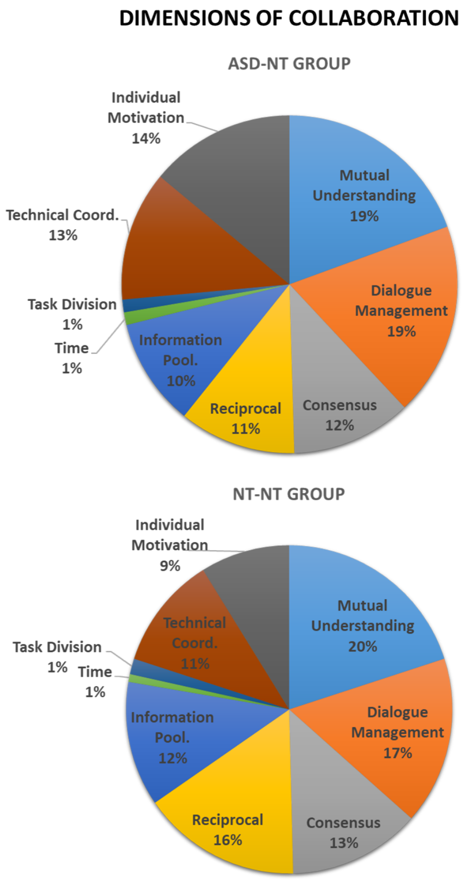

Fig. 7. Dimensions of collaboration pattern in ASD-NT group and NT-NT group. Similar collaboration pattern overall except for Technical Coordination, Individual Motivation, and Reciprocal interactions.

dimensions of collaboration in the tasks, and 4) examine collaboration patterns in autistic individuals and neurotypical individuals. Based on the results and analysis presented in the previous section, we believe that the collaborative tasks designed for ViRCAS have the potential to motivate and encourage teamwork between autistic and neurotypical individuals. Our findings offer preliminary support that ViRCAS can assist individuals with ASD and without ASD in learning work-relevant teamwork skills, and capture multimodal data across tasks of varying difficulty levels using different input devices.

The use of multimodal data made it possible to provide quantitative measures of the different dimensions of collaboration. For example, we were able to use input device data to assess technical coordination. The same information may not be easily available from the transcribed speech or other data. Also, multimodal data provided us with quantitative measures of collaboration that human observers might not capture from observing the session or watching a video recording of the interaction such as gazed objects and manipulated objects, which can be difficult to capture by human observers but contain important information to represent collaborative actions. In general, multimodal data analysis can provide higher accuracy compared to unimodal analysis [56]. Currently, the simulator is comprised of three collaborative tasks across varying vocational domains, but it is not limited to these tasks alone. Future work can evaluate additional task types to determine the relevance and performance of the system across different job-relevant teamwork scenarios.

Effective teamwork requires collaborative effort by individuals to work together to achieve a common goal [32]. We embedded 9 dimensions of collaboration that can represent teamwork in our collaborative tasks. Participants in both ASD-NT group and NT-NT group showed similar patterns of collaboration, which could indicate that the tasks met the universal design principles where participants exhibit similar responses even though they have different abilities. Task division and time management dimensions had less than 10% occurrence in both groups. It is possible that the tasks were designed in a structured manner that offered fewer opportunities for the participants to divide them, and that participants were afforded ample time to not seek time management strategies. In the future we plan to modify the tasks to provide more opportunities for task division and time management.

We found that the difficulty levels increased participants' collaboration, as can be seen by the higher number of utterances and back-and forth conversations in the Hard level as compared to the Easy level. We did not compare the utterances in the Tutorials as they focused on task familiarization. Increased difficulty levels increased the ambiguity and task interdependence which motivated participants to ask more questions or describe the task more to each other to be able to proceed with the task. The utterance analysis showed increased collaborative effort for all participants working together. This observation is consistent with other studies that suggest more word usage can influence collaborative learning and learning gains [57]. Paired with dialogue acts analysis, we found that the increased in number of utterances were related to the tasks. We found statistically significant increases for utterances labeled as Acknowledgement ('Acks), Describe ('Desc'), Negative sentiment ('Neg'), Questions ('Ques'), 'Inform', and 'Read', which are all task-oriented conversations. For the Furniture Assembly task, we observed that utterances labeled as 'Read' were fewer in the more difficult level because the written instructions were removed in the more difficult level, while utterances labeled as Describe ('Desc') increased significantly across all participant groups because participants needed to describe what they see without the instructions. We also found that ASD-NT group used more description ('Desc') utterances compared to the NT-NT group, which could indicate that pairs in the ASD-NT group needed additional explanation and description when performing the task together. Participants in the ASD-NT group also had fewer utterances compared to the participants in the NT-NT group, which could be directly related to social communication difficulties faced by individuals with ASD [18]. However, reduced spoken words did not negatively impact teamwork performance as they were all able to work together to finish the tasks.

In the gaze analysis, we found that both autistic and NT individuals have similar gaze duration in the PC Assembly and Furniture Assembly tasks. However, for Fulfillment Center, NT participants spent 3 times longer looking at the crates compared to participants with ASD. Because the Fulfillment Center task involves driving a forklift, it is possible that aspects of driving (e.g., familiarity and comfort with driving, ability to focus on driving-relevant stimuli) artificially impacted performance of autistic participants. Studies related to driving in young adults with ASD have reported reduced gaze awareness on targeted areas [58], and altered gaze pattern compared to control groups [59], which is consistent with our findings for the Fulfillment Center task. Therefore, future work may consider assessing driving familiarity and comfort when utilizing a task with a driving component.

Our work emphasized input of stakeholders in preliminary task design and in offering feedback on the developed system. Both autistic and NT participants felt comfortable and confident when performing the collaborative tasks. However, as the tasks became more challenging, NT participants felt less comfortable than the autistic participants. It is unclear if this was related to the tasks themselves or to the complexities of social interaction with autistic partners.

As for the input device preference, the haptic device was the least favorite device compared to the other devices, which indicated that ease-of-use was more important to the participants than immersive interaction since participants were least familiar with the use of haptic device. However, existing studies that explored the use of haptic devices in VR-based interactions have shown that with practice, haptic devices can be well accepted by participants [27], [33].

VI. Conclusion

Teamwork skills are one of the core skills sought by employers as they can contribute to improved productivity and workplace performance [14]. However, differences in communication and social interaction skills in autistic adults relative to their neurotypical colleagues can lead to poor teamwork performance, thus limiting employment opportunities for autistic individuals where a high level of teamwork is required [13]. Motivated by this, we designed a novel collaborative activities simulator within CVE, ViRCAS, to support teamwork skills practice for both autistic and neurotypical adults. Results from a feasibility study with 12 participant pairs indicated preliminary acceptance of ViRCAS, a positive impact of the collaborative tasks on supported teamwork skills practice for both autistic and neurotypical individuals, and promising potential to quantitatively assess collaboration through multimodal data analysis.

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

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

References

[1] L. Kenny, C. Hattersley, B. Molins, C. Buckley, C. Povey, and E. Pellicano, "Which terms should be used to describe autism? perspectives from the uk autism community," *Autism*, vol. 20, no. 4, pp. 442–462, 2016.

[2] M. Guha, "Diagnostic and statistical manual of mental disorders: Dsm-5," *Reference Reviews*, 2014.

[3] M. J. Maenner, K. A. Shaw, A. V. Bakian, D. A. Bilder, M. S. Durkin, A. Esler, S. M. Furnier, L. Hallas, J. Hall-Lande, A. Hudson *et al.*, "Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2018," *MMWR Surveillance Summaries*, vol. 70, no. 11, p. 1, 2021.

[4] P. M. Dietz, C. E. Rose, D. McArthur, and M. Maenner, "National and state estimates of adults with autism spectrum disorder," *Journal of Autism and Developmental Disorders*, vol. 50, no. 12, pp. 4258–4266, 2020.

[5] "Growing numbers of young adults on the autism spectrum," 2019. [Online]. Available: https://drexel.edu/autismoutcomes/blog/overview/2019/June/Growing-numbers-of-young-adults-on-the-autism-spectrum/

[6] A. M. Roux, J. E. Rast, and P. T. Shattuck, "State-level variation in vocational rehabilitation service use and related outcomes among transition-age youth on the autism spectrum," *Journal of Autism and Developmental Disorders*, vol. 50, no. 7, pp. 2449–2461, 2020.

[7] D. Hendricks, "Employment and adults with autism spectrum disorders: Challenges and strategies for success," *Journal of vocational rehabilitation*, vol. 32, no. 2, pp. 125–134, 2010.

[8] A. Karpur, V. Vasudevan, A. Lello, T. W. Frazier, and A. Shih, "Food insecurity in the households of children with autism spectrum disorders and intellectual disabilities in the united states: Analysis of the national survey of children's health data 2016–2018," *Autism*, vol. 25, no. 8, pp. 2400–2411, 2021.

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

[10] S. Baron-Cohen, E. Ashwin, C. Ashwin, T. Tavassoli, and B. Chakrabarti, "Talent in autism: hyper-systemizing, hyper-attention to detail and sensory hypersensitivity," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1522, pp. 1377–1383, 2009.

[11] F. Happé and U. Frith, "The weak coherence account: detail-focused cognitive style in autism spectrum disorders," *Journal of autism and developmental disorders*, vol. 36, no. 1, pp. 5–25, 2006.

[12] J. C. Kirchner and I. Dziobek, "Toward the successful employment of adults with autism: a first analysis of special interests and factors deemed important for vocational performance," *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology*, vol. 2, no. 2, pp. 77–85, 2013.

[13] M. Waisman-Nitzan, N. Schreuer, and E. Gal, "Person, environment, and occupation characteristics: What predicts work performance of employees with autism?" *Research in Autism Spectrum Disorders*, vol. 78, p. 101643, 2020.

[14] J. B. Schmutz, L. L. Meier, and T. Manser, "How effective is teamwork really? the relationship between teamwork and performance in healthcare teams: a systematic review and meta-analysis," *BMJ open*, vol. 9, no. 9, p. e028280, 2019.

[15] E. Walsh, J. Holloway, and H. Lydon, "An evaluation of a social skills intervention for adults with autism spectrum disorder and intellectual disabilities preparing for employment in ireland: A pilot study," *Journal of Autism and Developmental Disorders*, vol. 48, no. 5, pp. 1727–1741, 2018.

[16] "Neurodiversity hiring: Global diversity and inclusion at microsoft." [Online]. Available: https://www.microsoft.com/en-us/diversity/inside-microsoft/cross-disability/neurodiversityhiring

[17] Specialisterne, "Specialsterne: Assessment." [Online]. Available: https://www.specialisterneni.com/about-us/assessment/

[18] W. Chen, "Multitouch tabletop technology for people with autism spectrum disorder: A review of the literature," *Procedia Computer Science*, vol. 14, pp. 198–207, 2012.

[19] V. Bernard-Opitz, N. Sriram, and S. Nakhoda-Sapuan, "Enhancing social problem solving in children with autism and normal children through computer-assisted instruction," *Journal of autism and developmental disorders*, vol. 31, no. 4, pp. 377–384, 2001.

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

[21] E. Salas, N. J. Cooke, and M. A. Rosen, "On teams, teamwork, and team performance: Discoveries and developments," *Human factors*, vol. 50, no. 3, pp. 540–547, 2008.

[22] K. Dillenburger, H.-R. Röttgers, K. Dounavi, C. Sparkman, M. Keenan, B. Thyer, and C. Nikopoulos, "Multidisciplinary teamwork in autism: Can one size fit all?" *The Educational and Developmental Psychologist*, vol. 31, no. 2, pp. 97–112, 2014.

[23] F. Lateef, "Simulation-based learning: Just like the real thing," *Journal of Emergencies, Trauma and Shock*, vol. 3, no. 4, p. 348, 2010.

[24] S. Parsons and P. Mitchell, "The potential of virtual reality in social skills training for people with autistic spectrum disorders," *Journal of intellectual disability research*, vol. 46, no. 5, pp. 430–443, 2002.

[25] L. Bozgeyikli, E. Bozgeyikli, A. Raij, R. Alqasemi, S. Katkoori, and R. Dubey, "Vocational rehabilitation of individuals with autism spectrum disorder with virtual reality," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 10, no. 2, pp. 1–25, 2017.

[26] "21st century skills and the workplace." [Online]. Available: https://www.gyli.org/wp-content/uploads/2014/02/21st_century_skills_Gallup.pdf

[27] E. Almaguer and S. Yasmin, "A haptic virtual kitchen for the cognitive empowerment of children with autism spectrum disorder," in *International Conference on Human-Computer Interaction*. Springer, 2019, pp. 137–142.

[28] E. Bekele, Z. Zheng, A. Swanson, J. Crittendon, Z. Warren, and N. Sarkar, "Understanding how adolescents with autism respond to facial expressions in virtual reality environments," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 4, pp. 711–720, 2013.

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

[30] K. Lee, "The future of learning and training in augmented reality." *InSight: A Journal of Scholarly Teaching*, vol. 7, pp. 31–42, 2012.

[31] C. Williams, B. Wright, G. Callaghan, and B. Coughlan, "Do children with autism learn to read more readily by computer assisted instruction or traditional book methods? a pilot study," *Autism*, vol. 6, no. 1, pp. 71–91, 2002.

[32] J. E. Driskell, E. Salas, and T. Driskell, "Foundations of teamwork and collaboration." *American Psychologist*, vol. 73, no. 4, p. 334, 2018.

[33] H. Zhao, A. Zaini Amat, M. Migovich, A. Swanson, A. S. Weitlauf, Z. Warren, and N. Sarkar, "Inc-hg: An intelligent collaborative haptic-gripper virtual reality system," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 15, no. 1, pp. 1–23, 2022.

[34] M. R. Kandalaft, N. Didehbani, D. C. Krawczyk, T. T. Allen, and S. B. Chapman, "Virtual reality social cognition training for young adults with high-functioning autism," *Journal of autism and developmental disorders*, vol. 43, no. 1, pp. 34–44, 2013.

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

[36] E. Salas, K. A. Wilson, C. E. Murphy, H. King, and M. Salisbury, "Communicating, coordinating, and cooperating when lives depend on it: tips for teamwork," *The Joint Commission Journal on Quality and Patient Safety*, vol. 34, no. 6, pp. 333–341, 2008.

[37] S. Wallace, S. Parsons, and A. Bailey, "Self-reported sense of presence and responses to social stimuli by adolescents with autism spectrum disorder in a collaborative virtual reality environment," *Journal of Intellectual & Developmental Disability*, vol. 42, no. 2, pp. 131–141, 2017. [Online]. Available: https://doi.org/10.3109/13668250.2016.1234032

[38] S. Parsons, "Learning to work together: Designing a multi-user virtual reality game for social collaboration and perspective-taking for children with autism," *International Journal of Child-Computer Interaction*, vol. 6, pp. 28–38, 2015.

[39] Y. Song, L.-P. Morency, and R. Davis, "Multimodal human behavior analysis: Learning correlation and interaction across modalities," ser. ICMI '12. New York, NY, USA: Association for Computing Machinery, 2012. [Online]. Available: https://doi.org/10.1145/2388676.2388684

[40] A. Meier, H. Spada, and N. Rummel, "A rating scheme for assessing the quality of computer-supported collaboration processes," *International Journal of Computer-Supported Collaborative Learning*, vol. 2, no. 1, pp. 63–86, 2007.

[41] A. Z. Amat, M. Breen, S. Hunt, D. Wilson, Y. Khaliq, N. Byrnes, D. J. Cox, S. Czarnecki, C. L. Justice, D. A. Kennedy, T. C. Lotivio, H. K. McGee, D. M. Reckers, J. W. Wade, M. Sarkar, and N. Sarkar, "Collaborative virtual environment to encourage teamwork in autistic adults in workplace settings," in *Universal Access in Human-Computer Interaction. Design Methods and User Experience*, M. Antona and C. Stephanidis, Eds. Cham: Springer International Publishing, 2021, pp. 339–348.

[42] C. Solomon, "Autism and employment: Implications for employers and adults with asd," *Journal of autism and developmental disorders*, vol. 50, no. 11, pp. 4209–4217, 2020.

[43] T. Grandin, "Choosing the right job for people with autism or asperger's syndrome," 1999. [Online]. Available: https://www.iidc.indiana.edu/irca/articles/choosing-the-right-job-for-people-with-autism-or-aspergers-syndrome.html

[44] M. F. S. M.S., "Maximizing usability: The principles of universal design," *Assistive Technology*, vol. 10, no. 1, pp. 4–12, 1998, pMID: 10181150. [Online]. Available: https://doi.org/10.1080/10400435.1998.10131955

[45] "Webcams." [Online]. Available: https://www.logitech.com/en-us/products/webcams.html

[46] "Touch haptic device - 3d systems." [Online]. Available: https://www.3dsystems.com/haptics-devices/touch

[47] U. Technologies. [Online]. Available: https://unity.com/

[48] "Mirror networking – open source networking for unity." [Online]. Available: https://mirror-networking.com/

[49] [Online]. Available: https://webrtc.org/

[50] "Speech to text – audio to text translation: Microsoft azure." [Online]. Available: https://azure.microsoft.com/en-us/services/cognitive-services/speech-to-text/overview

[51] "Recognize speech from a microphone." [Online]. Available: https://docs.microsoft.com/en-us/azure/cognitive-services/speech-service/how-to-recognize-speech?pivots=programming-language-csharp

[52] "Celebrating 20 years - global leader in eye tracking," Apr 2015. [Online]. Available: https://www.tobii.com/

[53] A. Gibaldi, M. Vanegas, P. J. Bex, and G. Maiello, "Evaluation of the tobii eyex eye tracking controller and matlab toolkit for research," *Behavior Research Methods*, vol. 49, pp. 923 – 946, 2017.

[54] "Unity sdk - tobii developer zone." [Online]. Available: Unity SDK - Tobii Developer Zone

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

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

[57] N. Webb, M. Hepple, and Y. Wilks, "Dialogue act classification based on intra-utterance features," 01 2005.

[58] B. Reimer, R. Fried, B. Mehler, G. Joshi, A. Bolfek, K. M. Godfrey, N. Zhao, R. Goldin, and J. Biederman, "Brief report: Examining driving behavior in young adults with high functioning autism spectrum disorders: A pilot study using a driving simulation paradigm," *Journal of autism and developmental disorders*, vol. 43, no. 9, pp. 2211–2217, 2013.

[59] D. Bian, J. W. Wade, L. Zhang, E. Bekele, A. Swanson, J. A. Crittendon, M. Sarkar, Z. Warren, and N. Sarkar, "A novel virtual reality driving environment for autism intervention," in *Universal Access in Human-Computer Interaction. User and Context Diversity*, C. Stephanidis and M. Antona, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 474–483.