Design of an Intelligent Agent to Measure Collaboration and Verbal-Communication Skills of Children with Autism Spectrum Disorder in Collaborative Puzzle Games

Lian Zhang, Ashwaq Z. Amat, Huan Zhao, Amy Swanson, Amy Weitlauf, Zachary Warren, and Nilanjan Sarkar, *Senior Member, IEEE*

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

***Index Terms*—Adaptive and intelligent educational systems, autism spectrum disorder, collaborative learning, learning environments.**

I. Introduction

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

Computer-assisted interventions may offer an alternative intervention and assessment modality with reduced costs of care [8]. Many children with ASD have a natural affinity for computer-controlled environments [9] and exhibit a high level of engagement within these systems [10]. In addition, computer systems can provide controllable, replicable, and safe environments for children with ASD to practice social communication skills. As such, various kinds of computer-mediated intervention systems have been developed in order to understand and enhance the social communication skills of children with ASD [10], [11]. Among these are collaborative game-based interventions, which usually target two users' ability to convey information to one another (communicate) and to work together to achieve a common goal (collaborate) [12].

Collaborative games poses several advantages relative to

Manuscript received August 12, 2018; revised March 13, 2019; accepted March 6, 2020. This work was supported in part by the National Institute of Health Grants 1R01MH091102-01A1 and 1R21MH111548-01, and National Science Foundation Grant 0967170 and the Hobbs Society Grant from the Vanderbilt Kennedy Center. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

L. Zhang, A.Z. Amat and H. Zhao are with Department of Electrical Engineering and Computer Science and Robotics and Autonomous Systems Laboratory, Vanderbilt University, Nashville, TN, USA. (e-mail: lian.zhang@vanderbilt.edu; ashwaq.zaini.amat.haji.anwar@vanderbilt.edu; huan.zhao@vanderbilt.edu).

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

A. Weitlauf and Z. Warren are with Department of Pediatrics, Vanderbilt University Medical Center, Treatment and Research Institute for Autism Spectrum Disorders, Nashville, TN, USA. (e-mail: amy.s.weitlauf@vumc.org; zachary.e.warren@vumc.org).

N. Sarkar is with Department of Electrical Engineering and Computer Science, Department of Mechanical Engineering and Robotics and Autonomous Systems Laboratory, Vanderbilt University, Olin Hall Room 101, 2400 Highland Avenue, Nashville, TN, USA. 37212. (e-mail: nilanjan.sarkar@eanderbilt.edu).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TLT.2020.3029223, IEEE Transactions on Learning Technologies

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

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

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

A. Intelligent Agents with Conversation Capabilities

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

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

B. Intelligent Agents with Conversation Capabilities for Game Playing

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

Only a few intelligent agents with conversation capabilities have been designed to support and measure flexible conversations within the collaborative game domain. Cuayáhuitl and colleagues designed an artificial intelligent agent that can play a strategic board game, called Settlers of Catan [32]. In the board game, players can offer resources to other players and they can also reply to offers made by other players. Their study focused on applying a Deep Reinforcement Learning (DRL) method to train conversational skills of the

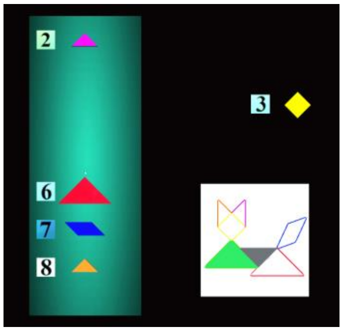


Fig. 1. An example of the collaborative puzzle games.

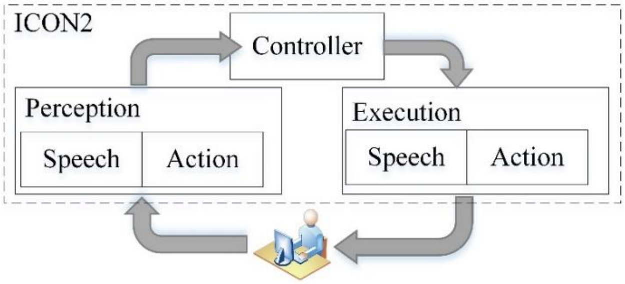


Fig. 2. Overall view of ICON2.

agent. Results of the study indicated that the DRL method significantly outperformed several other methods, including random, rule-based, and supervised methods, in training the agent's conversational skills. Kulms and colleagues designed an intelligent agent that could conduct text-based conversation as well as play a collaborative puzzle game [33]. In the collaborative puzzle game, the agent works together with a human to place differently shaped blocks in three steps:

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

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

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 evaluate their communication skills. In section II, we will first present the collaborative puzzle games, which were designed to encourage communication and collaborative interactions between children with ASD and the intelligent agent. The details of the game design are omitted here since they can be found in our previous publication [34]. In section III, the intelligent agent is described in detail with emphasis on its dialogue manager component. Section IV provides information about the experimental protocol and participants of two case studies. Results and discussions are presented in section V. Section VI shows the limitations and future works.

II. Collaborative Puzzle Games

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

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

III. Intelligent Agent

A. *Overall Description and Architecture*

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

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

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

1) Human user can ask ICON2, "What color is puzzle 5?" or

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

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

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

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

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

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

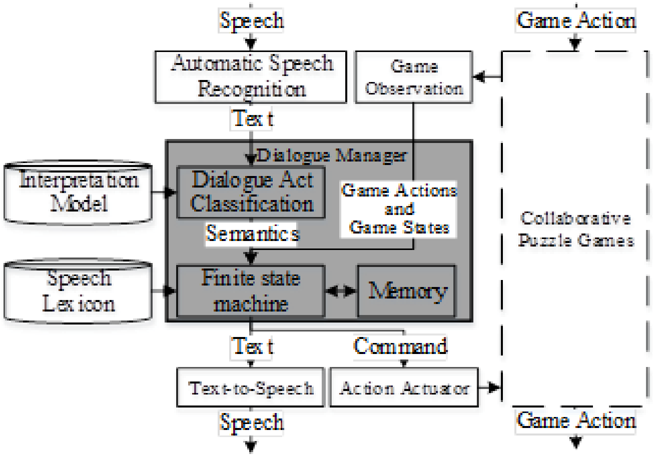


Fig. 3. Architecture of ICON2.

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

B. *Dialogue Manager*

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

TABLE I
DIALOGUE ACT CLASSES AND DESCRIPTIONS

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

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

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

An object indicates a specific puzzle piece targeted by an intention. Possible values of the object can be any of the seven puzzle pieces or empty. In order to simulate the real users' collaboration and verbal-communication behaviors, ICON2 must be able to 1) understand a human's intention; 2) find a targeted object; and 3) generate appropriate speech and game-related responses.

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

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

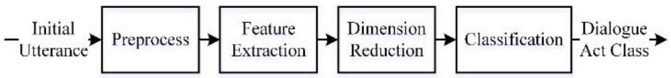


Fig. 4. The process of the online classification.

In order to fulfill these requirements, we developed the communication and game-playing behaviors of ICON2 in three steps: i) understanding human spoken natural language and collecting game-related inputs, ii) detecting intention and object from the speech and game-related inputs, and iii) generating speech and game-related responses. These three steps together could enable both communication and game-playing capabilities. The speech and game-related inputs gathered in the first step were not only important for ICON2 to communicate and play games but also useful for ICON2 to evaluate skills of its human-partner. In the third step, some rules were included so that ICON2 can also initiate conversations in addition to responding to the users. In what follows, we describe the first step in section II-B-1 and section II-B-2, the second step in section II-B-3, and the third step in section II-B-4.

1) *Language understanding*

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

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

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

2) *Game-related inputs*

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

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

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

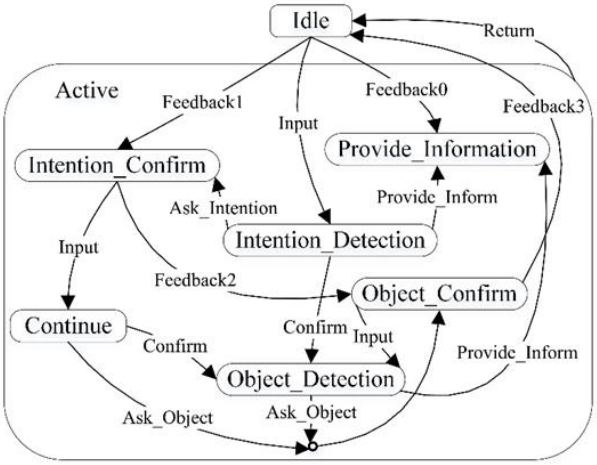


Fig. 5. Finite state machine in the Dialogue Manager Module.

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

3) *Intention and object detection*

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

The logic for intention detection can be summarized using the structure shown in Fig. 6. The tree-structure simplifies the intention-detection procedure by dividing it into multiple steps. In the first step, ICON2 detects out-of-domain utterances based on a rule: if an utterance has out-of-domain slot words, the utterance is an out-of-domain utterance. As mentioned in the

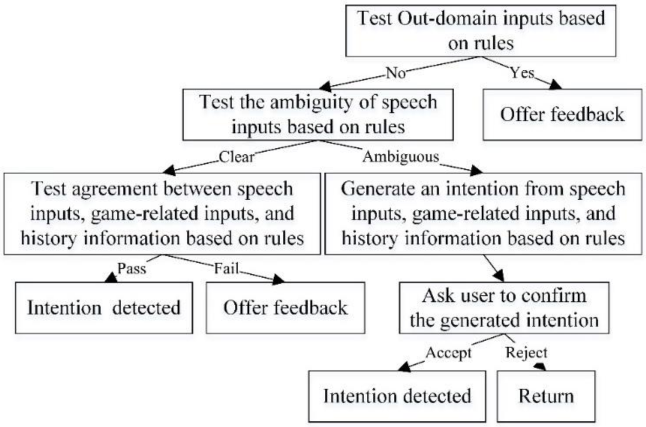

Test Out-domain inputs based on rules
No
Yes
Test the ambiguity of speech inputs based on rules
Offer feedback
Clear
Ambiguous
Test agreement between speech inputs, game-related inputs, and history information based on rules
Generate an intention from speech inputs, game-related inputs, and history information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

Fig. 6. The logic for intention detection.

introduction section, existing spoken dialogue systems are usually designed to operate over a limited and definite domain [52]. To ensure satisfactory user experience, sthe spoken dialogue system must be able to detect out-of-domain (OOD) utterances, and provide feedback to the user when OOD utterances are detected. Previous literature has applied classification methods to explicitly model OOD utterances for OOD detection [53]. However, collecting enough training data to model OOD utterances is time-consuming and laborious. Given the limited training data samples of this study, it is hard to create an OOD model with acceptable accuracy. Therefore, we used a rule-based method to detect OOD utterances in the current study. This method has been proven to be useful in the system, as discussed in the results session. Other advanced OOD detection methods will be explored in our system in the future.

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

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

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

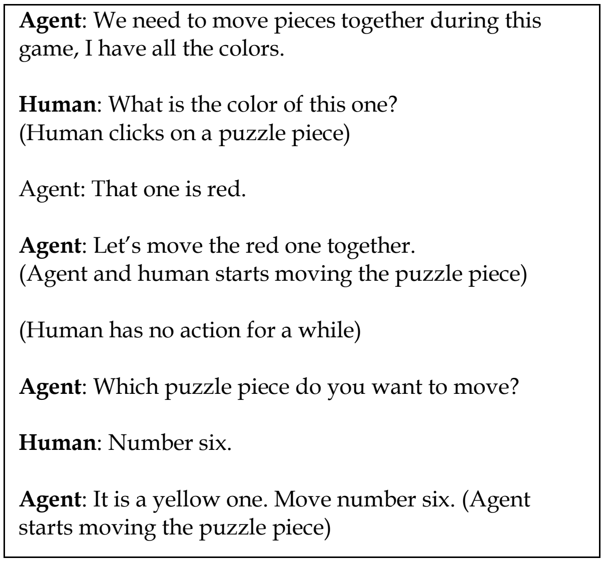
4) *Response generation*

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

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

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

1) “What is the color?”
2) “Could you tell me the color?”



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

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

Agent: That one is red.

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

(Human has no action for a while)

Agent: Which puzzle piece do you want to move?

Human: Number six.

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

Fig. 7. A sample dialogue (All game actions are shown in parentheses).

3) “Is it red or green?”

A sample dialogue is shown in Fig. 7.

IV. Preliminary Studies

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

A. *Collaborative Puzzle Games*

In Study I, we used seven collaborative puzzle games developed in our previous study [34]. The variation in these games was implemented by manipulating two game features: 1) who can move the puzzle pieces; and 2) who can see the colors of the puzzle pieces. The characteristics of the seven collaborative puzzle games are shown in Table II. Take

TABLE II
CHARACTERISTICS OF COLLABORATIVE PUZZLE GAMES IN PRELIMINARY STUDY I

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

Game_11 for example, where both users could see all the colors of the puzzle pieces, and they needed to take turns moving the pieces one by one. And for Game_15, only the human user could move the puzzle pieces, but ICON2 had the color information for the puzzle pieces. This forced the human user to ask ICON2 for the color information before choosing the puzzle piece to move to the target.

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

TABLE III
CHARACTERISTICS OF COLLABORATIVE PUZZLE GAMES IN PRELIMINARY STUDY II

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

B. *Participants and Experimental Procedure*

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

1) *Study I*

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

TABLE IV
THE CHARACTERISTICS OF THE FIVE PARTICIPANTS IN STUDY I

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)

2) *Study II*

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

TABLE V
THE CHARACTERISTICS OF THE FIVE PARTICIPANTS IN STUDY II

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

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

V. Skill Measurements Procedure

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

A. *System-Generated Features*

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

B. *Skill Measurements*

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

VI. Results

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

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

A. *Results of Study I*

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

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

TABLE VI
Automatically Generated Verbal-Communication and Task-Performance Features

Index	Feature	Description
1	Word frequency	How many words a user speaks per minute
2	Request color frequency	How many times per minute a user asks color information
3	Provide frequency	How many times per minute a user provides game information
4	Direct movement frequency	How many times per minute a user directs movements
5	Acknowledge frequency	How many utterances belong to acknowledgements
6	Request object frequency	How many times per minute a user asks for objects
7	Sentence frequency	How many utterances a user speaks in a minute
8	Success frequency	How many puzzle pieces have been successfully moved to the target area
9	Failure frequency	How many times a user fails in moving puzzle pieces
10	Collaboration time	The time duration of puzzle pieces being moved by two users simultaneously in a minute
11	Dragging time	The total time duration of a user dragging puzzle pieces
12	Collaborative movement ratio	The ratio of collaboration time and dragging time

with an average score of 4.4/5, where 1 meant dislike very much and 5 meant like very much.

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

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

TABLE VII
SURVEY RESULTS

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

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

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

We defined that all the utterances generated by the feedback events of the FSM were "initiations," and all the other utterances were "replies". In this study, ICON2 generated 161 initiations and 190 replies. 82.93% of the 161 initiations were labeled as appropriate initiations; while 89.33% of the 190 replies were labeled as appropriate replies. These results indicate that ICON2 demonstrates potential to communicate with the participants with ASD when they play puzzle games. Note that the accuracy of appropriate replies (89.33%) is much higher than the accuracy of dialogue act classification (67.47%), which suggests that ICON2 could reply appropriately even when it misunderstood a human's language. This was because ICON2 could reduce language ambiguity by combining the language with game-related inputs, as discussed in section III-B-2.

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

TABLE VIII
DIALOGUE ACT CLASSIFICATION RESULTS IN STUDY I

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

TABLE IX
DIALOGUE ACT CLASSIFICATION ACCURACIES IN STUDY II

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

the games.

B. *Results of Study II*

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

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

TABLE X
ERROR RATE OF EACH SYSTEM-GENERATED FEATURE

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

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

TABLE XI
ACCURACY TO ASSESS BOTH COMMUNICATION AND COLLABORATION SKILLS BASED ON THE SYSTEM-GENERATED FEATURES

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

VII. CONCLUSIONS

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

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

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

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

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

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

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

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

Fourth, the system-generated features were limited as well. We only explored 12 features for the measurements in the current study. Human behaviors, such as eye gaze, body language, and facial expression, could also provide important information in peer-mediated interactions. However, features to represent these behaviors have not been explored in this study.

In the future, these features will be captured using eye gaze recognition, gesture recognition, and emotion recognition in order to understand the non-verbal communications.

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[18] A. Ben-Sasson, L. Lamash, and E. Gal, "To enforce or not to enforce? The use of collaborative interfaces to promote social skills in children with high functioning autism spectrum disorder," *Autism*, vol. 17, pp. 608-622, 2013, doi: https://doi.org/10.1177/1362361312451526.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

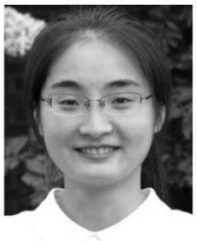
[61] R. Yaghoubzadeh, K. Pitsch, and S. Kopp, "Adaptive grounding and dialogue management for autonomous conversational assistants for elderly users," in *Int. Conf. Intell. Virtual Agents*, 2015, pp. 28-38, doi: 10.1007/978-3-319-21996-7_3.

Lian Zhang has just completed her Ph.D. degree from the Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, USA. She received her M.S. degree from the School of Automation Science and Electrical Engineering, in 2012, from Beihang University, Beijing, China.

She is now working as a Software Engineer with Facebook in California. Her research interests include affective computing, virtual reality, machine learning, and human-computer interaction.

Ashwaq Zaini Amat earned her Bachelor of Engineering in Electrical and Biomedical Engineering from Vanderbilt University, Nashville, TN, USA in 2009. She went on to complete her MS degree in Embedded Systems Engineering at University of Leeds, UK in 2014.

From 2009 to 2013, she was working as an IT consultant. She is now in her third year of her Ph.D. studies in electrical engineering. Her research interests include machine learning, human computer interaction and robotics.



Huan Zhao had recently completed her Ph.D. degree in electrical engineering at Vanderbilt University Nashville, TN, USA. She received her B.S. degree in automation from Xi’an Jiaotong University, Xi’an, China, in 2012, and the M.S. degree in electrical engineering from Vanderbilt University in 2016.

From 2012 to 2014, she did research at Institute of Artificial Intelligence and Robotics at Xi’an Jiaotong University.

Ms. Zhao is currently working at Facebook in California. Her research interests lie in the design and development of systems for special needs using the technologies of virtual reality, human-machine interaction, and robotics.

Amy R. Swanson received her M.A. degree in social science from the University of Chicago, Chicago, IL,USA, in 2006.

She is currently a Clinical/Translational Research Coordinator with Vanderbilt Kennedy Center’s Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), Nashville, TN, USA.

Amy S. Weitlauf received the Ph.D. degree in psychology from Vanderbilt University, Nashville, TN,USA, in 2011.

She is a licensed Clinical Psychologist with the Vanderbilt Kennedy Center’s Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), Nashville, and is an Assistant Professor of Pediatrics with Vanderbilt University Medical Center, Nashville.

Zachary E. Warren received the Ph.D. degree from the University of Miami, Miami, FL, USA, in 2005.

He is currently an Associate Professor of Pediatrics and Psychiatry with Vanderbilt University, Nashville, TN, USA.

Dr. Warren is the Director of the Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), Vanderbilt Kennedy Center, Nashville.

Nilanjan Sarkar (S’92–M’93–SM’04) received the Ph.D. degree in mechanical engineering and applied mechanics from the University of Pennsylvania, Philadelphia, PA, USA, in 1993.

After a postdoctoral fellowship at Queen’s University, Canada, he joined the University of Hawaii as an assistant professor in mechanical engineering. His current research interests include human–robot interaction, affective computing, dynamics, and control.

In 2000, Dr. Sarkar joined Vanderbilt University, Nashville, TN, USA, where he is currently the department chair and professor of mechanical engineering and a professor of electrical engineering and computer science. He is a Fellow of the ASME. He served as an associate editor for the IEEE Transactions on Robotics.