2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Assisting Group Discussions Using Desktop Robot Haru

Fei Tang[1], Chuanxiong Zheng[1], Hongqi Yu[1], Lei Zhang[1], Eric Nichols[2], Randy Gomez[2] and Guangliang Li[1*]

Abstract— **Socially assistive robots are potentially to be integrated with human daily lives in the near future, and expected to be able to improve group dynamics when interacting with groups of people in social settings. In this paper, we developed a system with desktop robot Haru to assist group discussions. The system consists of three modules: a dialogue assistance module which facilitates Haru to speak to users and answer questions in a free way; a dialogue balance module to encourage participation of users in the discussion with verbal behaviors; an autonomous gazing behavior module trained via deep reinforcement learning in simulation and deployed on physical Haru in reality, which can show politeness during group discussion, e.g., gazing to the speaking member, looking to the middle when both members are talking or silent, looking at the least spoken person when encouraging her. Results of user study with 40 subjects show the significant effectiveness of our system in assisting group discussion.**

I. INTRODUCTION

Social robots are potentially to be integrated with human daily lives in the near future and provide social, behavioral, emotional and cognitive support to people with diverse characteristics and needs [1]. In both our personal and professional lives, many problems we confront are solved in small groups [2], e.g., family conversation on financial problems, work plan discussion, group discussion for assignments in classroom etc. In these cases, socially assistive robots are expected to be able to improve group dynamics when interacting with groups of people in social settings. With their embodied nature, social robots inherently can interact with people in physical spaces, which creates opportunities to assist and improve human-human interaction [3].

Recent work has discovered that robot behaviors that can positively shape specific social dynamics in groups and teams of people. For example, Birmingham et al. [3] proposed a novel framework for robot mediation to study trust dynamics in group discussion using a Nao robot. However, in their work, a wizard controlled the robot's head direction and gaze direction to look at the speaking member of the group, and controlled the timing of the robot's speech. In addition, the robot balanced the number of questions posed to the participant, not trying to encourage members' participation based on detected group activity. Weldon et al. [4] studied how a robot can balance attention in a collaborative learning environment and found a non-verbal gaze mediating behavior in the robot led to increased gaze change frequency among participants as well as more time spent mirroring the robot's gaze. Instead of using humanoid robot, Tennent et al. [2] proposed to use Micbot—a peripheral robotic object, to promote participant engagement and problem solving performance with non-verbal implicit behaviors, such as facing to the speaking person and leaning the body of the microphone towards the participant who spoke least. In their work, the non-verbal behaviors are performed every fixed time interval and data from the microphone array was programmed to control the robot to move to the correct location, not used to learn autonomous behaviors. In addition, the robot only used non-verbal behaviors and cannot take turn to directly speak to the user. Shamekhi and Bickmore [5] developed a multi-modal robotic meeting facilitator using multi-modal sensor inputs (such as user gaze, speech, prosody, and proxemics) from group members to enforce meeting structure, promote time management, balance group participation in small group decision-making meetings. Sebo et al. [6] investigated the effects of verbal support from a robot (e.g., "good idea Salim," "yeah") on human team members' interaction and found increased verbal participation after receiving targeted support from the robot. Cumbal et al. [7] studied how encouraging vocal and non-vocal backchannels (e.g.,"uh huh") can stimulate speaking participation in a game-based multi-party interaction with a Furhat robot.

(a) (b)

Fig. 1: Examples of Haru's behaviors in our study. (a) Autonomous gazing at the speaking user; (b) Actively participating group discussion using our dialogue assistance module by integrating ChatGPT.

Currently, most research primarily used programmed or wizard controlled non-verbal robot behavior (e.g., gazing) and simple verbal support behavior (e.g., "good idea Salim," "yeah") to study group dynamics and increase participant engagement, and the robot did not take turn to speak to the group members. In this paper, we used animated desktop robot Haru and investigated how Haru's learned autonomous

[1]Faculty of Information Science and Engineering, Ocean University of China, Songling Road 238, 266100, Qingdao, Shandong, China. guangliangli@ouc.edu.cn

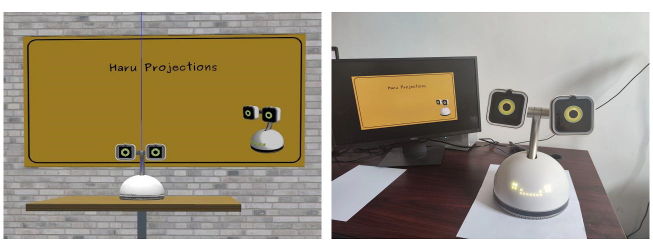
[2]Honda Research Institute Japan Co., Ltd, Wako, Japan. {e.nichols,r.gomez}@jp.honda-ri.com

[*] Corresponding author. This work was supported by Natural Science Foundation of China (grant No. 51809246), Qingdao Natural Science Foundation (grant No. 23-2-1-153-zyyd-jch) and Honda Research Institute Japan Co., Ltd.

non-verbal gazing behavior together with the encouraging verbal behavior affect the group discussion.

To this end, we developed a system with Haru consisting of three modules. A dialogue assistance module was designed by integrating autonomous speech recognition (ASR) with ChatGPT and text-to-speech (TTS), which facilitates Haru to speak to users and answer questions in a free way. A dialogue balance module was used to encourage participation of users in the discussion with verbal behaviors via the dialogue assistance module. Lastly, an autonomous gazing behavior module was trained via deep reinforcement learning in simulation and deployed on physical Haru in reality, which can show politeness during group discussion, e.g., gazing to the speaking member, looking to the middle when both members are talking or silent, looking at the least spoken person when encouraging her. We conducted a with-subject user study with 40 participants to evaluate the effectiveness of our system. Results of user study show that the encouraging verbal-behavior via our dialogue balance module can significantly increase the perceived politeness of Haru by participants and quality of discussion compared to a voice assistant with robot body. In addition, Haru performing both non-verbal gazing behavior and verbal encouraging behavior via autonomous gazing behavior module and dialogue balance module can further improve Haru's effectiveness in assisting group discussion.

II. ROBOT PLATFORM



(a) Simulated Haru (b) Physical Haru

Fig. 2: Simulated Haru on the Gazebo platform (a) and physical Haru in the real world (b).

Haru is an experimental multi-modal desktop robot with rich emotional capabilities [8], which enable verbal and non-verbal interactions with human users. As a desktop social robot, Haru offers five degrees of freedom to control limb movement, allowing for base rotation, neck tilt, eye movement, eye rotation and eye tilt. A matrix of addressable LEDs is embedded in the robot's body, which is capable of various combinations to simulate realistic mouth movements. Haru's eyes are enclosed within a sturdy shell, featuring a 3-inch TFT display bordered by addressable LED strips [9]. These visual elements provide visualizations that represent the robot's eye expressions. Through the integration of these hardware components, Haru can effectively express emotions through speech, paralanguage, and body movements [10], [11], [12], [13], [14], [15]. Fig. 2 shows the simulated Haru on the Gazebo platform and physical Haru in the real world.

III. METHODOLOGY

Our objective is to facilitate Haru to assist group discussion by learning autonomous gazing behavior to address etiquette issues, answering questions, and maintaining a balanced participation in the dialogue. To achieve this, we designed a system with Haru composed of three modules: autonomous gazing behavior module, dialogue assistance module and dialogue balance module. The structure and workflow of our system are illustrated in Fig. 3.

A. Autonomous Gazing Behaviour Module

The Autonomous Gazing Behavior module aims to realise the politeness of Haru by learning autonomously during a multi-party conversation. For simplicity, we consider two users conversing with each other and Haru is taken as an assistant. In the future, we would like to expand to scenarios with more users.

Specifically, we employed the deep reinforcement learning soft actor-critic (SAC) [16] method as the learning algorithm for the Autonomous Gazing Behavioral module. Reinforcement learning is a paradigm where an agent learns to perform a sequential decision task task by interacting with the environment via trial and error [17], [18]. In the learning process, at each discrete time step t, an agent observes its current state s_t in the environment, selects an action a_t based on the state, and then the environment transitions to a new state s_{t+1}. The agent receives a reward r_{t+1} based on the tuple $[s_t, a_t, s_{t+1}]$, which will be used to update its policy π. $\pi(a|s) = p(a_t = a|s_t = s)$ represents how the agent selects an action a_t given a state s_t. This cycle repeats until the agent learns an optimal policy π which will receive most accumulated rewards.

1) State Representation: The state representation consists of the three-dimensional coordinate (x, y, z) of Haru and the two conversing human users in the Cartesian coordinate system, which is then transformed into a spherical coordinate system. In the coordinate system, Haru's base center is taken as the origin, Haru's initial eye direction as the 0 azimuth angle, and the tabletop on which Haru is located as the reference plane with an elevation angle of 90 degrees. In addition, whether the two conversing users are engaged in the conversation is also considered to represent Haru's state in part.

2) Action Space: We set the angle to move as the output action of our reinforcement learning model which will be used to directly control the robot motor. It's worth noting that when transferring the trained policy in simulation to physical Haru in reality, the motor rotation is limited a safe range to reduce the strain and potential burden on the motors.

3) Reward Function: We defined a reward function for Haru to learn autonomous gazing behavior, as below:

$$Reward = \begin{cases} 1 - \frac{2|\alpha - \alpha_n|}{\beta}, & \alpha_n - \beta \leq \alpha \leq \alpha_n + \beta. \\ -1, & else. \end{cases} \tag{1}$$

$$\beta = \frac{1}{2}\left|\alpha_2 - \alpha_1\right|, \tag{2}$$

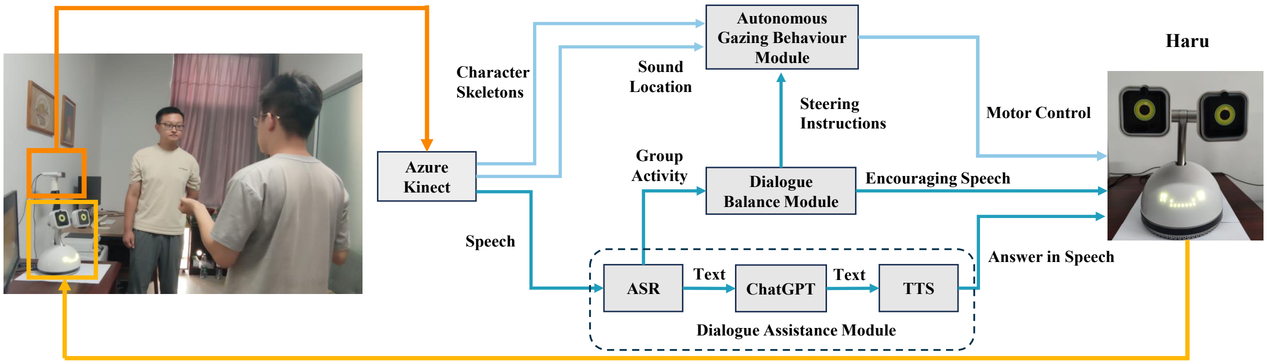


Fig. 3: Illustration of the structure and workflow of our system.

$$\alpha_n = \begin{cases} \alpha_1, & \textit{Only human user 1 is speaking}, \\ \alpha_2, & \textit{Only human user 2 is speaking}, \\ \alpha_0, & \textit{else}, \end{cases} \quad (3)$$

where α_n represents the optimal angle for Haru to perform at current step, α denotes Haru's current angle from the initial direction. The optimal angle α_n varies based on Haru's current state, which can be α_1, α_2, and α_0 respectively, as shown in Fig. 4(a). Specifically, when user 1 is speaking, $\alpha_n = \alpha_1$ — the angle from Haru's initial direction to user 1; when user 2 is speaking, $\alpha_n = \alpha_2$ — the angle from Haru's initial direction to user 2; when both users are silent or speaking simultaneously, $\alpha_n = \alpha_0$ — the angle from Haru's initial direction to the middle point of user 1 and 2. Fig. 4(b) shows the situation of human user 1 speaking where $\alpha_n = \alpha_1$, and the blue shaded area covering both sides of α_n is the reward interval. Haru will receive a reward that is proportional to the deviation from the optimal angle if α falls within reward interval, as in Equation 1. The rewards are normalized to be within the range of [-1, 1]. If α falls outside the reward interval, the reward received by Haru will always be -1.

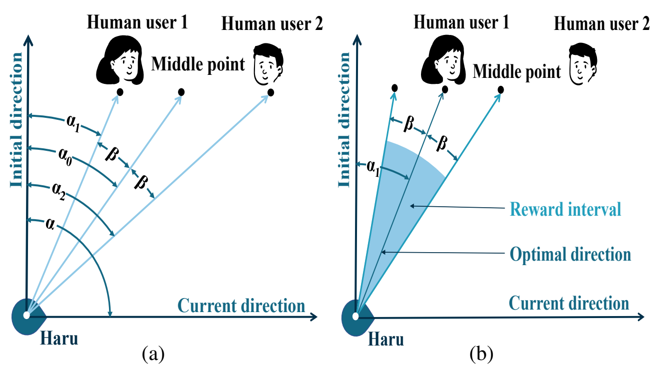


Fig. 4: (a) Haru's possible optimal angle α_n from the initial direction in different situations; (b) the reward interval in the situation of human user 1 speaking ($\alpha_n = \alpha_1$).

B. Dialogue Assistance Module

The dialogue assistance module primarily consists of three sub-modules: Autonomous Speech Recognition (ASR), Chat-GPT [19] and Text-to-Speech (TTS). The ASR sub-module is responsible for recognizing the speech and converting speech to text in real time through Google's API [20]. The recognized results are then sent to ChatGPT [21] to obtain answers to questions in the recognized speech content. After that, the obtained answers from ChatGPT are further converted to speech via the TTS sub-module and conveyed to the human user with the microphone on Haru, thus completing one round of dialogue with the human user.

C. Dialogue Balance Module

The Dialogue Balance module aims to maintain a balance of user's participation in the group dialogue by quantifying individuals' engagement. The engagement of each human user is measured as a "group activity" score from three perspectives: the duration of time she speaks, the number of times she participates in the dialogue, and the number of questions she asks Haru. For example, every time the user speaks for 0.5-2 seconds, she will receive 1 point for the "group activity" score. Additionally, for each question asked by the user to Haru, she will receive 80 points for her "group activity" score. Lastly, each participant is awarded 5 "group activity" points each time she participates in the group dialogue.

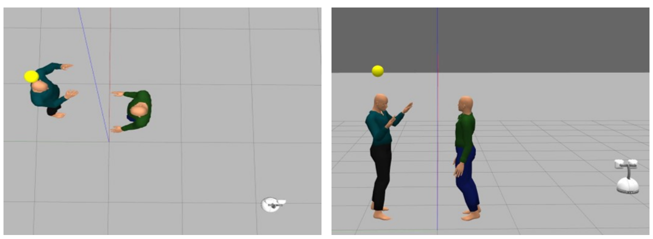
Haru will continuously monitor all participants' engagement and record their "group activity" scores during the group dialogue. Each time after Haru responds to a question raised by one human user or when the silence time of the group dialogue reaches a threshold, it will evaluate the participation of users by comparing their "group activity" scores. If the "group activity" score difference between users exceeds 100, Haru will directly communicate to the user with the lower score, aiming to encourage and enhance her participation in group conversations. For example, Haru will speak some open questions to raise the user's attention and motivate her to talk, such as "Hi, Bob! What are your thoughts?" Or " perhaps you can try answering the question from a different perspective?".

IV. EXPERIMENTS

As it usually takes a very long time for an agent to learn a good behavior policy via deep reinforcement learning and a considerable amount of interaction time can wear out the robot's hardware [22], [23], we trained Haru to learn the autonomous gazing behavior via deep reinforcement learning in our simulation platform and deployed the trained policy on the real Haru robot. Then, we tested Haru with transferred autonomous gazing behavior policy and our designed system in a real dialogue experiment with human users.

A. Autonomous Gazing Behavior Training in Simulation

We built a simulation environment for autonomous gazing behavior training with simulated Haru on Gazebo. In the simulation, two human users are engaged in a face-to-face conversation, while their body positions change slightly to simulate the real-life conversations of human users. To improve the policy's generalization, Haru was allowed to appear at different positions on either side of the two individuals to simulate different situations that might occur in a real-life conversation. Fig.5(a) and Fig.5(b) show one possible situation of Haru's position and the two conversing individuals from side view and top review respectively. In addition, a yellow ball was shown to indicate Haru's expected focal point.



(a) Top view (b) Side view

Fig. 5: Built simulation environment for autonomous gazing behavior training:(a) top view and (b) side view. The yellow ball was shown to indicate Haru's expected focal point.

We trained Haru via the SAC algorithm in the designed simulation environment. During the training process, Haru will receive rewards according to the designed reward function in Equation 1, ranging from -1 to 1. To encourage exploration, at the beginning of the training process, Haru was set to perform 10,000 random actions (i.e., steps) using a random policy instead of Haru's learned policy, and Haru's policy is not updated until the 1,000th step. After the 1,000th step, Haru's policy is updated every 50 steps until Haru finishes the 10,000 random actions. Then, Haru's policy will be used for action selection and trained for 200 iterations in total with each iteration consisting of 400 steps.

B. Real Dialogue Experiment With Human Users

Study Design: To assess the effectiveness of the three modules (i.e., Haru's behaviors) in our system in the group dialogue, we set up three conditions and conducted a real dialogue experiment with human users:

- Condition 1: only the Dialogue Assistant module (i.e., chatting behavior) is activated, which is like a simplified version of voice assistants like Apple's Siri [24], Amazon's Alexa [25], and Google's Google Assistant [26], but with a robot body;
- Condition 2: Both the Dialogue Assistant module and Dialogue Balance module (i.e., chatting and dialogue balancing) are activated on Haru;
- Condition 3: All three modules—Dialogue Assistant Module, Dialogue Balancing Module and Autonomous Gazing Behaviour Module (i.e., chatting, dialogue balancing and autonomous gazing), are activated. The Autonomous Gazing Behaviour Module works by transferring the trained simulation policy to physical Haru.

Procedure: There are two parts in our experiment. In the first part, we evaluated our system with human users in a group dialogue. 24 participants were recruited from a university campus, ranging in age from 20 to 25 years old. The mean age was 23.3 years old, with a standard deviation of 1.127. Among the participants, 54.17% were male, 33.33% were female, and 12.5% opted not to disclose their gender. All participants provided consent to participate in the study and were divided into 12 groups with each group consisting two participants: one is assumed to be a "talkative" student and the other to be a "shy" one. All participants took part in the three experimental conditions and the order of each group participating the three conditions was randomly assigned. The two participants were allowed to familiarize with each other for five minutes before the experiment started. In the experiment, the two participants of each group received a topic to talk from a staff with role of "teacher" and engaged in a discussion on the topic in front of a table, with Haru positioned on the table between them. The robot was regarded as a member of the group, and served as the group assistant. Azure Kinect was used to get the skeleton and sound location of users. A camera was mounted on a tripod to record the process of the dialogue. Then, the two participants filled out a questionnaire after finishing experiments in each condition. Participants were allowed to take a break for a few minutes between participating different experimental conditions.

The first questionnaire was used to evaluate the effectiveness of Haru and participants' experience in the group discussion. The effectiveness of Haru was evaluated from four perspectives: anthropomorphic, animacy, likability, and perceived intelligence, which were adopted from Godspeed's questionnaire [27]. In addition, a customized scale with four questions was administered to asses the participants' experience: "Does Haru's presence make you feel uneasy?" , "How satisfied are you with Haru?", "Would you like to continue using Haru for the next discussion?", and "Do you think Haru could be introduced into the classroom in the future?" All questions are on a 7-point Likert scale.

The second part of the experiment was set to evaluate the effectiveness of Haru and participants' discussion from a teacher's point of view. We recruited 16 participants to watch the recorded videos of the three experimental conditions in the first part. All participants are between 21 and 25 years old, with a mean age of 23.5 years old and a standard deviation of 1.212. Among these participants, 68.75% were male, 25% were female, and 6.25% chose not to disclose their gender. In the experiment, each participant was instructed to play the role of "teacher" and evaluate a group discussion involving two students and a robot assistant in a recorded video. Each participant watched all videos selected in random order from the three conditions, and was allowed to watch each video for multiple times. The participant was asked to fill out a second questionnaire after watching each video. The second questionnaire includes the same questions as the first one to evaluate the effectiveness of Haru from the four perspectives: anthropomorphic, animacy, likability, and perceived intelligence. In addition, the customized scale with four questions were adapted to asses the group members' performance from a teacher's point of view. Finally, both questionnaires concluded with an open-ended question: "What comments or suggestions do you have on the robot?".

V. RESULTS & DISCUSSION

A. Autonomous Gazing Behavior Learning

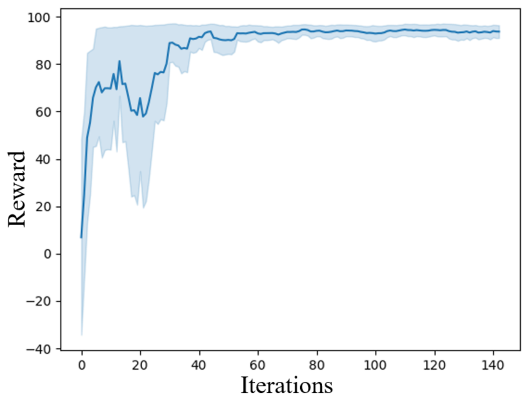
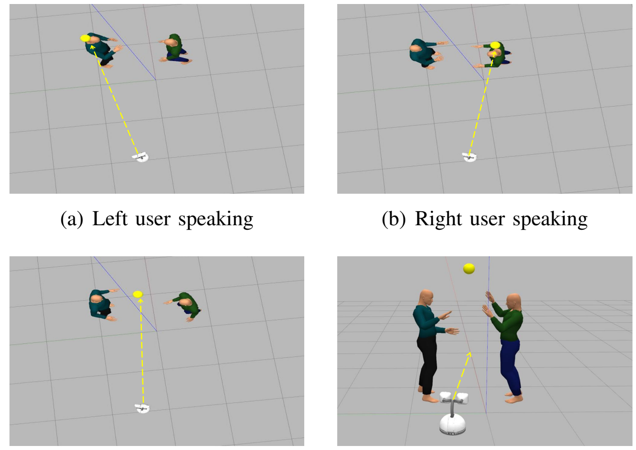


Fig. 6: Haru's autonomous gazing behavior learning curve in simulation.

Fig. 6 shows Haru's learning curve for autonomous gazing behavior trained in simulation, measured with cumulative reward received in one episode consisting of 100 steps according to the defined reward function in Section III A. The blue line represents Haru's mean performance by testing the learned policy for three trials and the shaded area indicates the maximum and minimum performance. From Fig. 6 we can see that, Haru can generally learn the autonomous gazing behavior in a few iterations' training. After about 40 iterations' training, Haru can already learn an optimal behavior. Fig. 7 shows some examples of Haru's learned optimal gazing behavior, with a yellow sphere as label on top of person to indicate where Haru should look and who is speaking. From Fig. 7 we can see that, Haru can direct its gaze to the speaking user and to the middle of users when both users are talking.

B. Haru's Performance in the Real Experiment

Fig. 8 shows examples of Haru's behaviors and performance in the experimental process of the three conditions in the user study. From Fig. 8 we can see that, in Condition 3 where all three modules are activated, Haru will shift its gaze to the talking participant and try to look at the shy participant (autonomous gazing behavior) and encourage her to speak with the Dialogue Balance module. While in Condition 2 with Autonomous Gazing Behavior module inactivated, Haru tried to improve the shy user's participation with the Dialogue Balance module, but did not look at her.



(a) Left user speaking (b) Right user speaking

(c) Both users speaking (top view) (d) Both users speaking (side view)

Fig. 7: Examples of Haru's learned optimal gazing behavior.

In Condition 1, the discussion time is shorter compared to Condition 2 and 3, since Haru did not try to balance users' participation in the dialogue and look at the talking user.

Haru's performance was also assessed from both the student's and teacher's point of views by comparing responses collected from the two questionnaires in the user study. 20 valid responses were collected from the first questionnaire (2 groups did not finish the experiment) and 16 from the second questionnaire. Cronbach's reliability analysis was conducted to analyze the reliability of data collected from both questionnaires. Results show that Cronbach's coefficient is higher than 0.80 on average for the first questionnaire and above 0.85 on average for the second questionnaire, demonstrating good internal consistency with response values for each participant across questions.

1) Student's Point of View: We did analysis of variance (ANOVA) analysis as well as post-hoc test — Fisher's Least Significant Difference test — on collected responses to questions in the first questionnaire (significance level: $p<0.01$), which was assessed by comparing mean scores rated by the 20 "students" in the questionnaire after the study from five perspectives: Anthropomorphism, Animacy, Likeability, Perceived Intelligence and Quality of Discussions, as shown in Fig. 9. Our ANOVA analysis shows that there is a significant difference between conditions across all perspectives ($p<0.01$) Post-hoc tests show that the effectiveness of Haru and the quatlity of students' discussion in Condition 3 were significantly better than those in Condition 2 and 1 ($p<0.01$ and $p<0.01$, respectively). Moreover, the effectiveness of Haru and quality of students' discussion in Condition 2 were also significantly better than those in Condition 1 ($p<0.01$).

2) Teacher's Point of View: We also did an analysis of variance (ANOVA) analysis as well as post-hoc test — Fisher's Least Significant Difference test — on collected responses to questions in the second questionnaire (significance level: $p<0.01$), which was assessed by comparing mean scores rated by the 16 'teachers' in the questionnaire after the study from the five perspectives: Anthropomorphism, Animacy, Likeability, Perceived Intelligence and Quality of Discussions, as shown in Fig. 10. Our ANOVA analysis

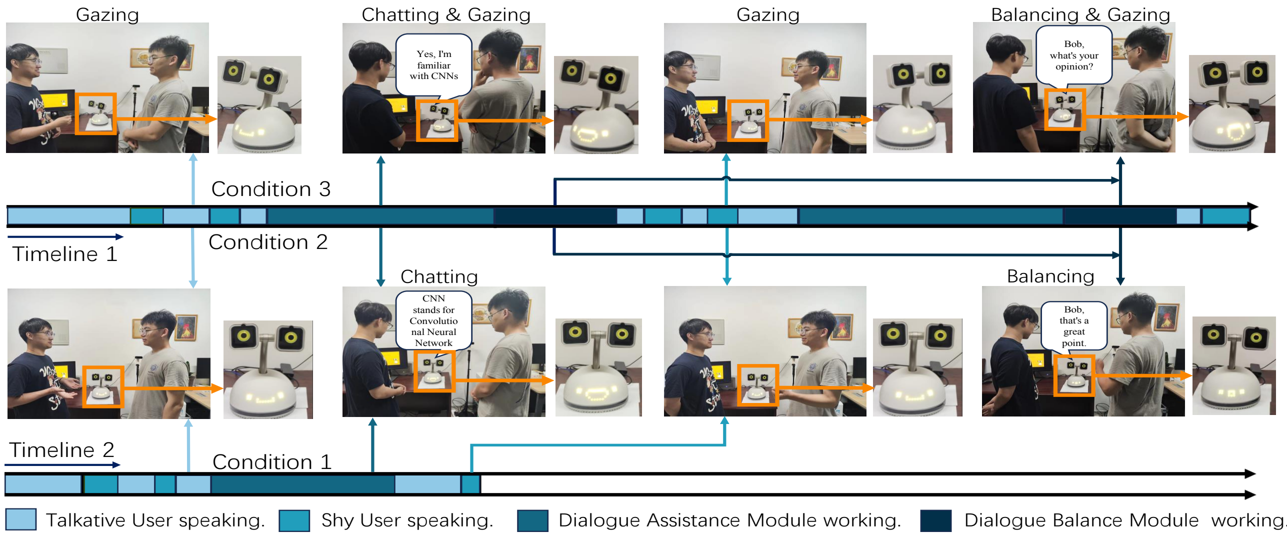


Fig. 8: Examples of Haru's behaviors and performance in the three experimental conditions of the user study. Timeline 1 shows Haru's behaviors in Condition 3 (autonomous gazing, chatting and dialogue balancing) and 2 (chatting and dialogue balancing), which have similar length of time in the experiment. Timeline 2 shows the only chatting behavior for Haru in Condition 1.

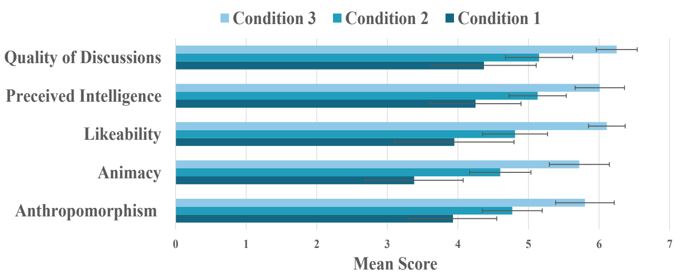


Fig. 9: Mean scores of the effectiveness of Haru and students' discussion rated by the 20 participants ('students') in the first questionnaire after the study in the three conditions. Note: black bars represent standard deviation.

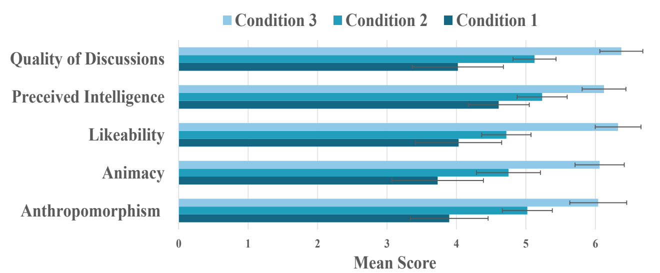


Fig. 10: Mean scores of the effectiveness of Haru and students' discussion rated by the 16 participants ('teachers') in the second questionnaire after the study in the three conditions. Note: black bars represent standard deviation.

shows that there is a significant difference between conditions across all perspectives ($p < 0.01$). Post-hoc tests show that the effectiveness of Haru and quality of students' discussion in Condition 3 were also significantly better than those in Condition 2 and 1 ($p<0.01$ and $p<0.01$, respectively). Moreover, the effectiveness of Haru and quality of students' discussion in Condition 2 were also significantly better than those in Condition 1 ($p<0.01$), except in terms of 'Animacy' and 'Perceived Intelligence' where the differences are close to significance ($p = 0.03$ and $p = 0.02$).

In addition, we received 12 responses to our open questions regarding the experiment and robot Haru. 8 of them mentioned that Haru's behaviors in Condition 1 and 2 are much less polite than those in Condition 3, and 7 of them felt satisfied with Haru's performance in Condition 3.

In summary, our analysis shows that Haru's dialogue balancing behavior (Condition 2) can significantly increase the effectiveness of Haru and quality of discussion in the group dialogue from both 'student's' and 'teacher's' point of view, compared to a chatting Haru (Condition 1). Moreover, activating both Dialogue Balance Module and Autonomous Gazing Behavior Module (Condition 3) can further increase the effectiveness of Haru and quality of discussion significantly.

VI. CONCLUSION

In this paper, we developed and implemented a system for assisting group discussion with social robot Haru. Our system consists of three modules: dialogue assistance module, dialogue balance module and autonomous gazing behavior module. The dialogue assistance module integrating autonomous speech recognition with text-to-speech and ChatGPT facilitates Haru to talk to users, while the dialogue balance module was used to balance participation of users in the discussion. The autonomous gazing behavior module was trained via deep reinforcement learning in simulation and deployed on physical Haru in reality. Our user study shows that our system can significantly increase the politeness of Haru and quality of discussion compared to a voice assistant with robot body (chatting Haru). In the future, we would like to extend our system to group dialogue with more than two users and evaluate the effectiveness of our system in a more realistic setting, e.g., a classroom with children.

ACKNOWLEDGMENT

We thank Paulo Alvit and Gonçalo Dias from IDMIND for their support and help with our experiment.

REFERENCES

[1] Chung Hyuk Park, Raquel Ros, Sonya S Kwak, Chien-Ming Huang, and Séverin Lemaignan. Towards real world impacts: Design, development, and deployment of social robots in the wild. *Frontiers in Robotics and AI*, 7:600830, 2020.

[2] Hamish Tennent, Solace Shen, and Malte Jung. Micbot: A peripheral robotic object to shape conversational dynamics and team performance. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 133–142. IEEE, 2019.

[3] Chris Birmingham, Zijian Hu, Kartik Mahajan, Eli Reber, and Maja J Matarić. Can i trust you? a user study of robot mediation of a support group. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, pages 8019–8026. IEEE, 2020.

[4] Catherine F Weldon, Sarah Gillet, Ronald Cumbal, and Iolanda Leite. Exploring non-verbal gaze behavior in groups mediated by an adaptive robot. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pages 357–361, 2021.

[5] Ameneh Shamekhi and Timothy Bickmore. A multimodal robot-driven meeting facilitation system for group decision-making sessions. In *Proceedings of International Conference on Multimodal Interaction (ICMI)*, pages 279–290, 2019.

[6] Sarah Sebo, Ling Liang Dong, Nicholas Chang, Michal Lewkowicz, Michael Schutzman, and Brian Scassellati. The influence of robot verbal support on human team members: Encouraging outgroup contributions and suppressing ingroup supportive behavior. *Frontiers in Psychology*, page 3584, 2020.

[7] Ronald Cumbal, Daniel Alexander Kazzi, Vincent Winberg, and Olov Engwall. Shaping unbalanced multi-party interactions through adaptive robot backchannels. In *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*, pages 1–7, 2022.

[8] Randy Gomez, Deborah Szapiro, Kerl Galindo, and Keisuke Nakamura. "haru": Hardware design of an experimental tabletop robot assistant. In *Procedings of the 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 233–240, 2018.

[9] Eleanor Sandry, Randy Gomez, and Keisuke Nakamura. Art, design and communication theory in creating the communicative social robot 'haru'. *Frontiers in Robotics and AI*, 8:577107, 2021.

[10] Eleanor Sandry. *Robots and communication*. Springer, 2015.

[11] Fernando Poyatos. The reality of multichannel verbal-nonverbal communication in simultaneous and consecutive interpretation. *Benjamins Translation Library*, 17:249–282, 1997.

[12] Yurii Vasylkiv, Zhen Ma, Guangliang Li, Eleanor Sandry, Heike Brock, Keisuke Nakamura, Irani Pourang, and Randy Gomez. Automating behavior selection for affective telepresence robot. In *Proceddings of IEEE International Conference on Robotics and Automation (ICRA)*, pages 2026–2032. IEEE, 2021.

[13] Yurii Vasylkiv, Zhen Ma, Guangliang Li, Heike Brock, Keisuke Nakamura, Irani Pourang, and Randy Gomezv. Shaping affective robot haru's reactive response. In *Proceedings of the 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 989–996. IEEE, 2021.

[14] Hui Wang, Jinying Lin, Zhen Ma, Yurii Vasylkiv, Heike Brock, Keisuke Nakamura, Randy Gomez, Bo He, and Guangliang Li. Affective behavior learning for social robot haru with implicit evaluative feedback. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3881–3888. IEEE, 2022.

[15] Hui Wang, Guodong Chen, Randy Gomez, Keisuke Nakamura, Bo He, and Guangliang Li. Shaping haru's affective behavior with valence and arousal based implicit facial feedback. In *Proceedings of the 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 769–776. IEEE, 2022.

[16] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 1861–1870. PMLR, 2018.

[17] R. Sutton and A. Barto. *Reinforcement learning: an introduction*. MIT Press, 1998.

[18] M. L. Littman. Reinforcement learning : A survey. *Journal of Artificial Intelligence Research*, 4, 1996.

[19] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.

[20] Google. Speech-to-Text. Retrieved from `https://cloud.google.com/speech-to-text/`, 2023.

[21] OpenAI. ChatGPT API. Retrieved from `https://openai.com/chatgpt`, 2021.

[22] Hao Ju, Rongshun Juan, Randy Gomez, Keisuke Nakamura, and Guangliang Li. Transferring policy of deep reinforcement learning from simulation to reality for robotics. *Nature Machine Intelligence*, pages 1077–1087, 2022.

[23] Jens Kober, J Andrew Bagnell, and Jan Peters. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11):1238–1274, 2013.

[24] Siri. `https://www.apple.com/ios/siri/`. Accessed: September 10, 2023.

[25] Alexa. `https://www.amazon.com/alexa`. Accessed: September 10, 2023.

[26] Google Assistant. `https://assistant.google.com`. Accessed: September 10, 2023.

[27] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1:71–81, 2009.