Impact of Trajectory Generation Methods on Viewer Perception of Robot Approaching Group Behaviors

Conference Paper · August 2020

DOI: 10.1109/RO-MAN47096.2020.9223584

4 authors:

Fangkai Yang
KTH Royal Institute of Technology

Wenjie Yin
KTH Royal Institute of Technology

Mårten Björkman
KTH Royal Institute of Technology

Christopher Edward Peters
KTH Royal Institute of Technology

Impact of Trajectory Generation Methods on Viewer Perception of Robot Approaching Group Behaviors

Fangkai Yang[1†], Wenjie Yin[2†], Mårten Björkman[2], Christopher Peters[1]

***Abstract*— Mobile robots that approach free-standing conversational groups to join them should behave in a safe and socially-acceptable way. Existing trajectory generation methods focus on collision avoidance with pedestrians, and the models that generate approach behaviors into groups are evaluated in simulation. However, it is challenging to generate approach and join trajectories that avoid collisions with group members while also ensuring that they do not invoke feelings of discomfort. In this paper, we conducted an experiment to examine the impact of three trajectory generation methods for a mobile robot to approach groups from multiple directions: a Wizard-of-Oz (WoZ) method, a procedural social-aware navigation model (PM) and a novel generative adversarial model imitating human approach behaviors (IL). Measures also compared two camera viewpoints and static versus quasi-dynamic groups. The latter refers to a group whose members change orientation and position throughout the approach task, even though the group entity remains static in the environment. This represents a more realistic but challenging scenario for the robot. We evaluate three methods with objective measurements and subjective measurements from viewer perception, and results show that WoZ and IL have comparable performance, and both perform better than PM under most conditions.**

I. INTRODUCTION

As mobile robots continue to have increased autonomy in human-robot interactions and are expected to work together with humans in teams, the ability to robustly approach and join groups, such as *free-standing conversational groups* [1], is fundamental. When doing so, it is vital to generate safe and socially-acceptable paths that avoid collisions with group members and do not make them feel uncomfortable, for example, by violating their personal space [2]. Recent work [3], [4], [5], [6] has conducted experiments involving generating safe and socially-acceptable paths. However, they have shortcomings that limit their utility. Especially, most research focuses on generating approach behaviors towards groups whose members are assumed to be totally static throughout the approach task. However, many situations involve groups that while static in the environment, have *quasi-dynamic* members, i.e., the positions and orientations of group members change over time as they make adjustments to account for a variety of factors, from slight weight-shifts to changes of formation due to a change in the focus of attention of the group or role of individuals within it [7]. This also relates to changes of position and orientation in order to make space for newcomers to join the group [8]. Another limitation of previous work is the lack of evaluation with human participants and comparison with other trajectory generation methods.

To study these limitations, we conduct an experiment to evaluate robot approaching group trajectories that are generated by three methods: a WoZ (Wizard-of-Oz) approach [9], a procedural model [10] and an imitation learning based model [11]. They are selected to represent manually controlled models, computational models, and machine learning models, due to their good performance (excluding WoZ) in respective domains. Our experiment is conducted in a motion capture lab to better capture the behaviors of group members that are used as input to three methods (Figure 1). The major contributions of the paper are summarized as follows:

- We conduct novel experiments to evaluate three trajectory generation methods in human-robot interactions from viewer perception in which robots approach to join free-standing conversation groups.
- We consider robot approaching behaviors under various experiment conditions, including group types, camera viewpoints, and approaching directions.

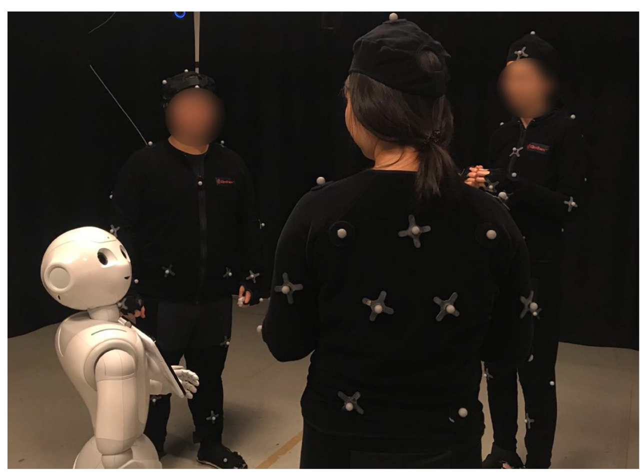

Fig. 1: The setup for our experiment in a motion capture lab where a Pepper robot approaches to join a free-standing group with three group members.

II. RELATED WORK

In this section, we present a summary of research on robot approaching group behaviors with related models to generate such behaviors and perception of robot movements.

A. Approaching Group Behaviors

Many studies have been carried out that specifically concern the approaching behaviors of robots into small free-

[1]Department of Computational Science and Technology, KTH Royal Institute of Technology, Stockholm, Sweden.

[2]Division of Robotics, Perception and Learning, KTH Royal Institute of Technology, Stockholm, Sweden.

[†]{fangkai, yinw}@kth.se.

standing conversational groups. As proposed by Kendon [12], the positions and orientations of individuals in such small groups are defined as the *F-formation* system. The central area within a group surrounded by group members, called *O-space* by Kendon, is an exclusive space that prevents robots from intruding into the group. For example, robots outside a group want to approach to join it, and they need to calculate a trajectory that does not intersect with the group's o-space.

Leveraging the F-formation system, Truong et al. [13] proposed a framework to enable a robot to approach a human group safely and socially. Claudio et al. [14] simulate approaching behaviors for virtual characters towards small groups. Samarakoon et al. [15] designed a rule-based method to replicate the natural approaching behaviors of humans. In recent work, Yang et al. [10] proposed a social-aware navigation method to navigate a robot to join a group in a socially-acceptable manner.

With the advent of deep learning, machine learning methods are being used to generate safe and social approaching group behaviors. Ramírez et al. [16] adopted inverse reinforcement learning, involving several participants demonstrating approaching behaviors for a robot to learn. Gao et al. [17] proposed a deep reinforcement learning model to generate robot approaching group behaviors. However, most methods involving groups consider them to be totally static, i.e., models assume that group members do not change body orientations and positions over the course of the approach. More recently, as pointed out in [7], individuals in groups are *quasi-dynamic*: even in static free-standing groups, individual members may routinely shift position and orientation. An approaching newcomer may, therefore, need to update their approach trajectory dynamically. By considering groups with quasi-dynamic individuals, Yang et al. [5] proposed a model based on Generative Adversarial Networks (GAN) to generate safe and socially acceptable trajectories into free-standing conversational groups that adapted to the position and orientation adjustments of group members.

B. Perception of Robot Movements

Existing research showed that many factors in robot approaching behaviors influence human perception. *Proxemics* is one vital factor that defines social distances between humans and robots. When robots approach a human subject or a group, they should not surprise the human subject and should keep a safe distance, while obeying social norms or being socially-acceptable [2]. Mead and Mataric [18] presented an experiment on how human participants perceive robots during interactions at various distances. Peters et al. [19] proposed comfortable proxemics when a virtual robot approaches a human user leveraging Hall's model [2]. Kuzuoka et al. [20] demonstrated a positional model based on proxemics that was able to reconfigure the F-formation.

Other works focus on the trajectories of the approach. Satake et al. [21] found that simply approaching a human subject by taking the shortest path is not enough to initiate human-robot interactions. Correspondingly, an analysis of approach behaviors towards groups by Yang et al. [22] showed that we do not necessarily approach a group directly, but take a comfortable path to the group.

Additionally, research has also focused on robot approach speeds, timings, and directions. Lohse et al. [23] found that inconstant robot speed during approaching behaviors increases a robot's legibility. Similarly, Henkel et al. [24] found that nonlinear approaching speed makes robots more acceptable. Huang et al. [25] found that robots with quick responses to user requests are perceived as more polite. Concerning approaching directions, studies show that people prefer to be approached within their field-of-view (FOV) rather than from directly behind [26]. Ball et al. [27] found that seated people feel least comfortable when the approaching robot cannot be seen. Moreover, there are works that focus on eye gaze [28], [29], facial/vocal emotions [30] in robot approaching behaviors and various human states, such as sitting [3], walking [31], standing [32].

III. METHOD

Our primary objective is to find a method to generate approaching group trajectories that are the most socially-acceptable to the group. To achieve this goal, we proposed three ways to generate the robot approaching behaviors.

A. WoZ

In the *WoZ (Wizard-of-Oz)* approach [9], the robot is teleoperated by a human operator (a researcher who is a trained operator). To better control the robot interacting with groups, both the camera view from the robot forehead camera from Choregraphe[1] and the reconstructed skeletons from Motive[2] (see Figure 2) are used to help the wizard perceive the environment and participants' full-body behaviors and reactions throughout the WoZ process.

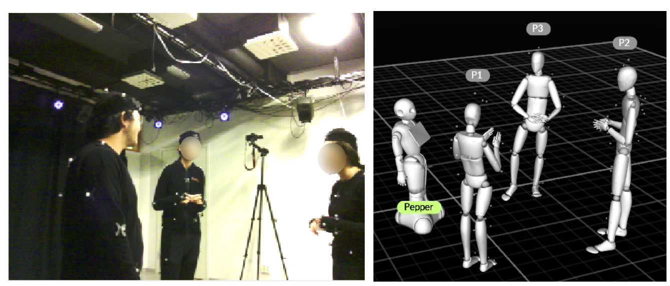

Fig. 2: Real-time views that help operator to control the robot. (Left) The camera view from the robot forehead camera. (Right) The reconstructed scene from Motive including three group members and the robot.

B. Procedural Model

We use the social-aware navigation method [10] as a procedural model to generate approaching group trajectories that are socially-acceptable and realistic. Additionally, Yang et al. [7] showed the social-aware navigation method is adopted to quasi-dynamic groups and outperforms other

[1] http://doc.aldebaran.com/2-4/software/choregraphe/index.html
[2] https://optitrack.com/products/motive/

state-of-the-art navigation methods, including A* [33] and SF (Social Forces) [34]. The procedural model is built upon a social-aware space [10] as shown in Figure 3, where a darker value area in the social-aware space means the robot walks in a lower speed and it will be harder to walk through. A fast marching method [35] was used to navigate the robot. It is an efficient method to track the motion of wavefronts. Due to the wave expansion properties, the path following the wavefront from the target point to the start point will be the fastest path, a path that is unique and complete.

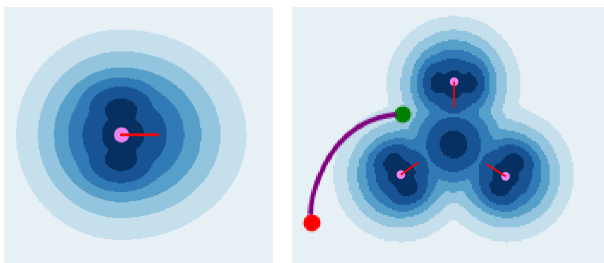

Fig. 3: The social-aware space of a standing agent facing right (left) and a free-standing conversational group with three group members (right) given a top-down view. The robot starts from the red point and approaches the group along the purple curve.

C. Imitation Learning Model

Imitation learning aims to learn a policy from expert observations. We propose to generate approaching group trajectories use a Generative Adversarial Imitation Learning (GAIL) [11] framework with a Group Behavior Recognition framework [22]. In our method, the policy generator π_θ generates state-action pairs (S × A) while the discriminator D_ω is trained to distinguish a trajectory generated by the policy generator π_θ from the expert observations π_E. The objective function is $\mathbb{E}_{\pi_\theta}[D_\omega(s,a)] - \mathbb{E}_{\pi_E}[D_\omega(s,a)]$.

In each training step, expert trajectories τ_E sampled from the CongreG8 dataset [22] are fed into the discriminator with mini-batch updates, together with sampled trajectories τ_θ generated by the policy generator π_θ. The discriminator is updated with the objective function, and its parameters ω_i are then clipped between (-0.01, 0.01). After updating the discriminator, the policy parameters θ are updated with Proximal Policy Optimization (PPO) [36].

We adopt a Group Behavior Recognition framework similar to the AG-GCN [22] in the policy generator and discriminator as the state encoder. The full-body markers of the three players are connected as skeleton graphs and fed into a Spatial Graph Convolutional Neural Network (S-GCN), which encodes the marker's spatial relationships into feature vectors. Past generated or expert trajectories are fed into an LSTM network, which encodes the past trajectories into feature vectors. On the group level, the feature vectors of the three players and the past trajectory are fed to the Group Graph Convolutional Neural Network (G-GCN). The steps involved in the imitation learning algorithm are summarized in Algorithm 6.

Algorithm 1 GAIL Algorithm for Robot Approaching Group Trajectory Generation

Input:
Input: player markers and expert trajectories
Initialize policy generator π_θ and discriminator D_ω
for $i = 0, 1, 2, \ldots$ **do**
  Sample expert τ_E and generated τ_θ trajectories
  Update discriminator parameters from ω_i to ω_{i+1} with gradient ascent on mini-batches with objective function: $\mathbb{E}_{\pi_\theta}[D_\omega(s,a)] - \mathbb{E}_{\pi_E}[D_\omega(s,a)]$
  Update policy parameters from θ_i to θ_{i+1} using the PPO rule.
end for

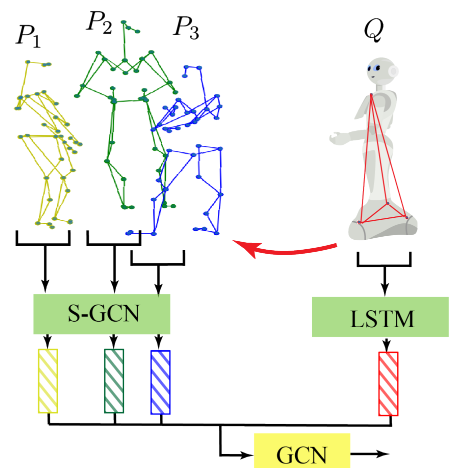

Fig. 4: The architecture of the state encoder

IV. Experiment

A. Experiment Design

To explore our research questions, we designed an exploratory scenario for human-robot interactions. In order to keep consistent with the dataset [8] used in imitation learning (Section III-C), a similar game *Who's the Spy* was used as the scenario (See Figure 5). This game involves three players in a group. In every game round, each player is given a card with a word on it. Among the three cards, only one card has a different word. When the game starts, the players take turns to describe the material properties of the word they have in hand. The robot acts as an adjudicator to identify the player who holds the different word cards, i.e., the spy. Once the robot established the identity of the spy, it approaches to join the group to inform group members of the outcome.

In this paper, we use video footage rather than a full live interaction. Live interaction trials are particularly challenging to develop when they involve complex human behaviors that need to be reliable and replicable for statistical analysis. For example, it is challenging to keep the group member behaviors to be consistent in various trails when the robot is controlled by the three methods (Section III). As suggested in [37], subjective ratings in live and remote (via video) trials are similar. Shinozawa et al. [38] found that human decision making depends more on interaction environment

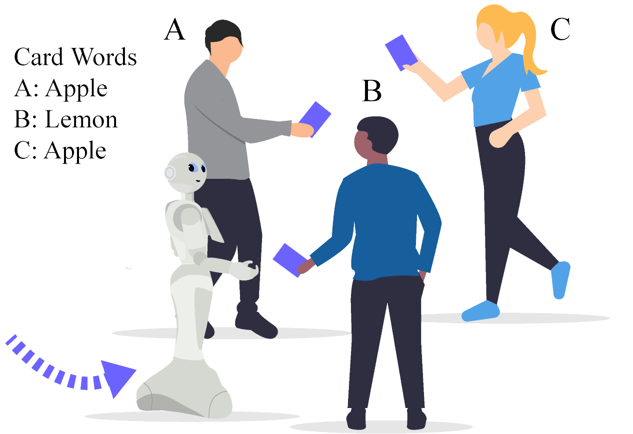

Fig. 5: The *Who's the Spy* scenario. Each of the three players has a card with a word on it, and only one word card is different. One player only knows the card on hand but does not know others. They take turns to describe the word, e.g., A: *It is a fruit.* B: *We add its juice to salmon.* C: *It's sweet if ripe.* The conversation goes on until the robot approaches to identify the spy, i.e., player B above, and all the players show the card on hand to confirm if the identification is correct.

and consistency. Additionally, Woods et al. [39] found that there are no significant findings for the subjective rating of the practicality for the robot approach direction task, and comfort levels between the live and video trial. We thus use video footage as stimuli in this paper as a way to reach a large number of participants. We record videos under various experimental conditions (see Section IV-B) performed by three players. The collected videos are shown to participants online for subjective assessment of approaching trajectories from viewer perception (see details in Section IV-E).

B. Experimental Conditions

1) Group Type: The experiment aims to explore the robot approaching group behaviors under three robot control methods. Most previous work [13], [17], [26], [27] focused on static groups where group members keep body positions and orientations unchanged in the human-robot interaction process. However, as proposed in [7], conversational groups are not always static but quasi-dynamic while the group as a whole is not moving, the group members may change position and orientation, and the robot needs to update its approaching trajectory. Yang et al. [8] confirmed the quasi-dynamic nature in conversational groups via data collection. In this paper, we thus explore robot approaching behaviors in static and quasi-dynamic groups.

In the static group, three players stand as a circle on three equally distributed positions (Figure 6). The group radius is 0.8 meters (the average value from the CongreG8 dataset [8]). The players stand on the initial location during the robot approaching procedure. On the other hand, the quasi-dynamic group has the same group radius.

Unlike the static group that has three possible entering points for the approaching robot, two players in the quasi-dynamic group stand closely at the start, which makes only two entering points available. As shown in Figure 7, while the robot approaches the group by following an initially planned path, player C notices the approaching robot and makes space for it. The robot thus changes the planned trajectory to join in the new entering point. In order to keep consistency in various quasi-dynamic trials, all the players stand on the same initial positions, i.e., two players stand closely initially to the opposite of the third player. After 2 seconds, one of the two closely standing players walks 0.7 meters aside and looks towards the robot to make space for it, showing the awareness of its presence.

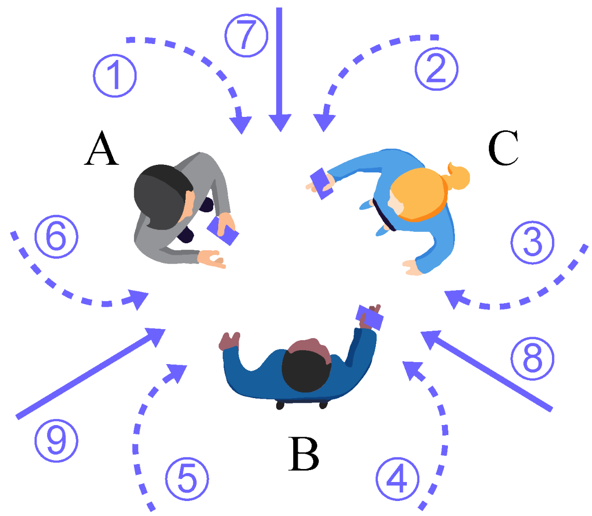

Fig. 6: A static group with three group members. The robot approaches from 9 directions including 6 direct approaching directions (1-6) and 3 indirect approaching directions (7-9).

2) Approaching Direction: Previous work [26], [27] evaluated the robot approaching directions towards either an individual person or two sitting persons. However, the approaching direction of a humanoid robot towards a conversational group remains unexplored. In the static group, the robot approaches from 9 directions, including direct and indirect ones (see Figure 6), starting 2.15 meters from the group center. On the other hand, the robot approaches and joins in the newly available entering point in the quasi-dynamic group (Figure 6). Note that a mirrored trial is performed where the robot starts from the back of player A, and at $T = T_1$ player A moves aside to make space. In this case, the robot initially plans to approach from the left and then changes to join in the newly opened position. We thus have two robot approaching directions in the quasi-dynamic group.

3) Camera Viewpoint: As suggested in [40], [41], [42], camera viewpoint influences human perception on agent behaviors that they are perceived as more salient in the egocentric view but with a clearer vision in the perspective view. We thus collected video stimuli from the egocentric view and the perspective view (see Figure 8).

C. Experiment Stimuli

Considering that the two group types have different robot approaching directions, we collected a different number of video clips in each group type scenario. In the static

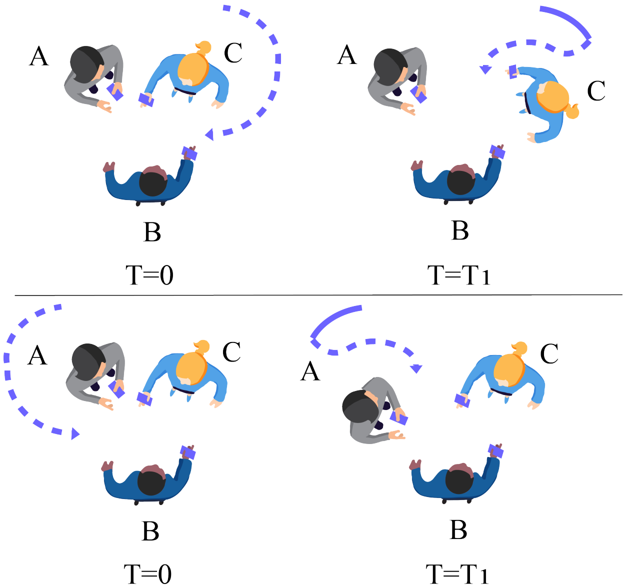

Fig. 7: A quasi-dynamic group with three group members. (Top) The robot started from the back of player C and planned a path (dash line) to approach and join the group. However, player C noticed the approaching robot and moved aside to make space for it at $T = T_1$, and the planned path changed. (Bottom) A mirrored case where the robot initially approached to the left and player A moved aside to make space.

Fig. 8: Collected videos in the egocentric view (left) and in the perspective view (right).

group, we collected 9 (approaching directions) $\times$ 2 (camera viewpoint) $\times$ 3 (robot control methods), i.e., 54 clips for the static group experiment scenario (27 clips in each camera viewpoint). On the other hand, in the quasi-dynamic group, we collected 2 (approaching directions) $\times$ 3 (robot control methods), i.e., 6 clips in the perspective view. However, since the quasi-dynamic group is asymmetric, we collected videos in the egocentric view of each player, and it results in 2 (approaching directions) $\times$ 3 (group members) $\times$ 3 (robot control methods), i.e., 18 egocentric clips. In total, we collected 78 clips that represent all condition combinations.

D. Experiment Apparatus

1) Motion Capture: As previously mentioned in Section III, both *Imitation Learning Model* and *Procedural Model* need real-time behavioral information of group members as inputs. We thus performed the experiment in a motion capture lab. The room has an approximate 5m $\times$ 5m $\times$ 3m active capture volume, which is equipped with a NaturalPoint Optitrack[3] system with 16 Prime 41 cameras. Each camera has a 4 mega-pixel resolutions with a frame rate of 120 fps. The motion of each group member was captured with a Motion Capture suit with 37 markers placed at respective anatomical locations of the body (see Figure 9 middle), and then the captured information is passed to our python program using NatNet SDK[4].

2) Robot: We used a physical Pepper[5] robot as the approaching group robot. Additionally, in order to enable a human operator to monitor the robot without actually seeing it (Section III-A), four motion capture markers were attached on its base and chest to track the motion, including its position and orientation via Motive (see Figure 9 left).

3) Video Recordings: We recorded robot approaching videos from both the egocentric view and the perspective view. The egocentric videos were recorded by attaching a GoPro Hero 5 camera to the upper-chest of one player (see Figure 9 middle), and the perspective videos were recorded with a Canon EOS 60D camera equipped with a Sigma 35mm f/1.4 camera lens on a tripod (see Figure 9 right).

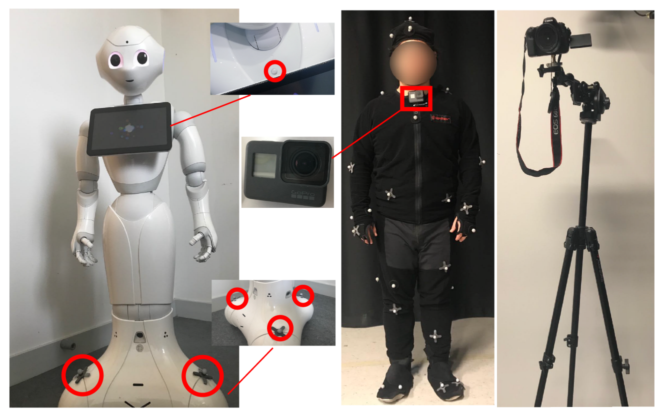

Fig. 9: The experiment apparatus: (Left) Pepper robot with four motion capture markers, (Middle) a player in a motion capture suit with 37 markers and a GoPro camera to take egocentric videos, (Right) Camera to take perspective videos.

E. Experiment Procedure

The experiment contains a video collection phase and a online survey phase.

1) Video Collection: First, an experimenter gave a brief introduction about the game to the players and informed them that they could stop their participation at any time. Prior to the start of the session, the players were asked to fill in consent forms. The experiments helped each player to select a Motion Capture suit with an appropriate size and to wear it. Then the videos were collected in the following phases:

Static Group Phase: The players were asked to keep standing on the fixed positions during the video recording phase, but they could talk and perform upper-body behaviors. The GoPro camera is fixed on the forehead of one player in this phase to collect egocentric video clips. Then the game

[3]https://optitrack.com/
[4]https://optitrack.com/products/natnet-sdk/
[5]https://www.softbankrobotics.com/emea/en/pepper

started, and the players took turns to describe word properties, and after 15 seconds, the robot approaches the group to identify the spy. One trial is thus collected via two cameras that result in an egocentric video and a perspective video. Then the next game started with new word cards, and the robot approached with another combination of approaching directions and robot control methods.

Quasi-dynamic Group Phase: The setup has discussed in Section IV-B. The player who moves aside to make space was asked to perform the same behaviors for other trials to keep consistency. Note that the GoPro camera was not attached to one player in this phase, but instead was switched among all players to capture asymmetric egocentric videos.

After the video collection, we found the egocentric camera could barely capture the robot approaching behaviors, and the player (who had the camera under the neck) realized the robot until the robot already joined the group. We thus remove clips that robots approach from back left and right, i.e., direction 4, 5, 8, 9 in Figure 6 if the egocentric camera is attached to Player B.

2) Online Survey: The online survey phase was divided into two sessions corresponding to the egocentric and perspective viewpoints, as shown in Figure 10. Within each session, the video clips were divided into two blocks concerning group types, and the ordering of video clips within each block was counter-balanced. Before each session, the participant received a training trial where a clip example with a questionnaire was presented. The data of the training session was not included in the analysis.

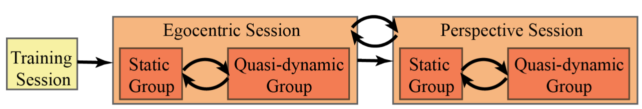

Fig. 10: The online survey procedure. Within each group block (red cube), the clips are counter-balanced. When it came to the next participant, the ordering of camera viewpoint sessions was switched as well as group blocks.

Twenty-seven participants (18M:9F) aged between 23-43 years old (M=28.1, SE=4.7) were recruited from the university locale to take the online survey. Most participants were not very familiar with robots. They were asked to watch the videos and answer three questions for each video. Specifically, participants were asked to rate how much they thought the robot approaching behaviors were *polite*, *human-like*, and *safe*, using a 1-7 Likert scale, where 1 means "not at all" and 7 means "very". These questions are designed to evaluate the robot approaching behaviors in three dimensions of social appropriateness, i.e., polite, social, and safe, as in [17], [43]. In the end, participants were asked to give feedback. Note that both players in the video collection and participants in the online survey were told that the robot was autonomous and lead them to believe that the robot was actively involved in the game during the group conversation.

V. Results

We present the results in both objective and subjective aspects for generated trajectories from WoZ, Procedural Model (PM), and Imitation Learning Model (IL) under various experimental conditions.

A. Objective Measurements

Two measurements, collision index (CI), and interaction index (II), are used to evaluate the social acceptableness of generated trajectories (see [13] for detailed definitions). CI is used to measure the physical safety, and II is used to evaluate the social interactions between the robot and groups. Note that WoZ and IL generate different trajectories even though the environment is the same. We thus randomly sampled one trajectory from WoZ and IL. Figure 11 shows the sampled trajectories that approach a static group (left) and a quasi-dynamic group (right) with related CI and II values. Trajectories generated by IL and WoZ have more deviated from the group so that they have lower risks of collision. PM makes abrupt turns when approaching a quasi-dynamic group, due to its costly computation that results in reaction delay. IL, however, adopts the quasi-dynamics by taking an early turn while keeping lower collision risks and higher group interactions.

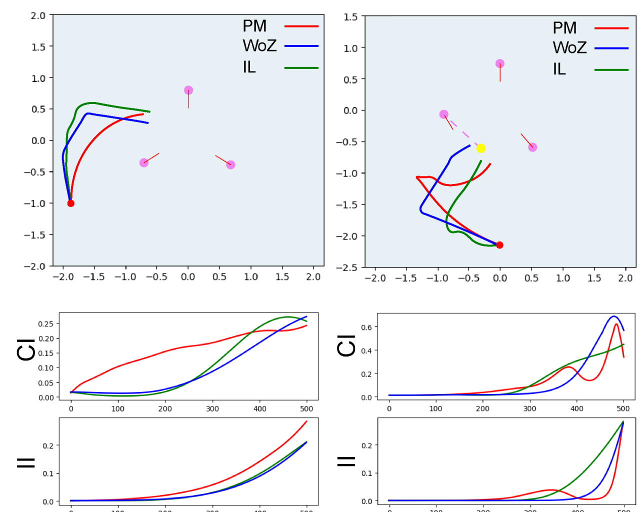

Fig. 11: Sampled trajectories (top) with related CI and II values (bottom) in a static group (left) and a quasi-dynamic group (right) where a group member moves aside from the initial position (yellow dot). The robot starts at the red dot and approaches groups (pink dots).

B. Subjective Measurements

The subjective measurement is done with the online survey analysis. The mean responses obtained in the different experimental settings were compared through a repeated measures Analysis of Variance (ANOVA) F-test. As the input factors are different in two viewpoints, we perform ANOVA tests for both egocentric session and perspective session. The egocentric session has factors including methods (Woz, PM or IL), approaching directions (left, front, right) and group type (static or quasi-dynamic), and the perspective session has methods (Woz, PM or IL), direction (directly or indirectly), and group type (static or quasi-dynamic) as input factors.

1) Egocentric Session: As shown in Figure 12 in a static group, people consider the approaching behaviors in the front direction as the least polite ($F(2,52) = 10.56, p < 0.01$) and the least sociable ($F(2,52) = 8.02, p < 0.01$). However, we found no significant difference in safety. One possible reason from the feedback is that people are fearful that the robot will collide even if the speed is imitated from human behaviors. IL has comparable scores as WoZ in politeness, sociality, and safety ($p's > 0.01$), excluding approaching in the front, and both are rated higher than the procedural model (PM).

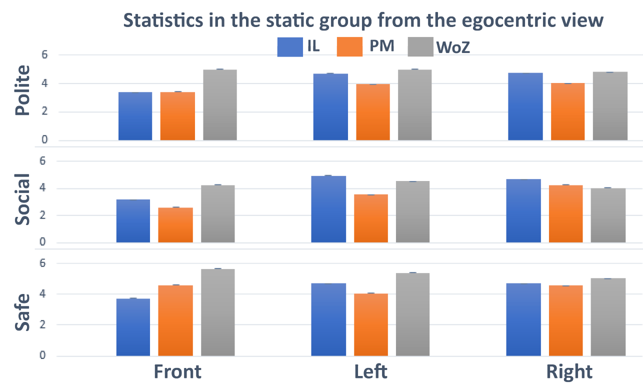

Fig. 12: Comparison of three methods in different approaching directions in the static group from the egocentric view.

In the quasi-dynamic group, as shown in Figure 13, there is no significant difference between IL and WoZ in politeness, sociality, and safety with $p's > 0.01$, and they both perform better than PM significantly.

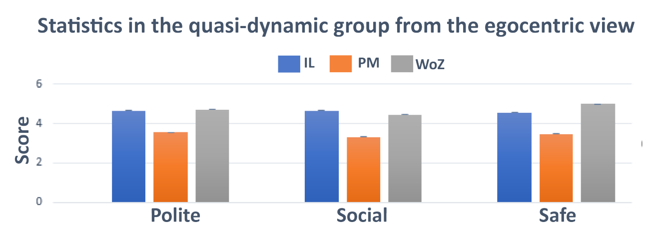

Fig. 13: Comparison of three methods in the quasi-dynamic group from the egocentric view.

2) Perspective Session: As shown in Figure 14, there is a significant effect of three methods in politeness ($F(2,52) = 5.14, p < 0.01$) and safety ($F(2,52) = 7.96, p < 0.01$), but not in sociality ($F(2,52) = 1.62, p > 0.01$). WoZ is rated as the highest in politeness and sociality significantly ($p's < 0.01$) if the robot approaches directly towards the group, but IL is rated as the highest in politeness and sociality when the robot approaches indirectly ($p's < 0.01$). Both WoZ and IL perform similarly in the safety dimension.

In the quasi-dynamic group (Figure 15), there is also a significant effect of methods ($F(2,52) = 7.81, p < 0.01$). Similar to the conclusion in the egocentric view, IL and WoZ are rated higher than PM in all three social appropriateness dimensions, and IL is comparable with WoZ.

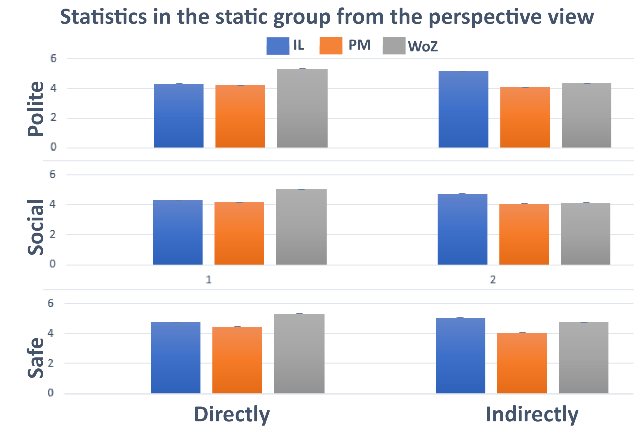

Fig. 14: Comparison of three methods in the static group from the perspective view.

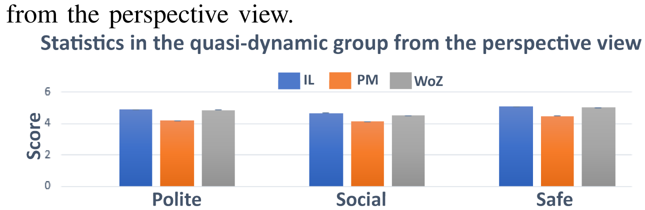

Fig. 15: Comparison of three methods in the quasi-dynamic group from the perspective view.

VI. CONCLUSIONS

In this paper, we conduct an experiment to evaluate three methods that generate robots approaching group trajectories under various experimental conditions. The imitation learning model has comparable performance with WoZ, and they both outperform the procedural model in objective measurements (with a lower risk of collision and higher group interaction) and subjective measurements from viewer perception. This offers a way to increase autonomy in mobile robots interacting with groups. It also raise a question to find a better solution to control mobile robot than imitating human behaviors. In the future, we will perform a similar IL model to output full-body behaviors and transfer imitated behaviors to mobile robots. We will experiment with groups in different formations to improve the adaptability of trajectory generation methods. Moreover, we will extract full-body behaviors from videos as input to adopt the imitation learning model in a general scenario without a motion capture setup.

ACKNOWLEDGEMENTS

This research has received funding from the European Union‘s Horizon 2020 research and innovation program under grant agreement n. 824160 (EnTimeMent).

REFERENCES

[1] X. Alameda-Pineda, Y. Yan, E. Ricci, O. Lanz, and N. Sebe, “Analyzing free-standing conversational groups: A multimodal approach,” in *Proceedings of the 23rd ACM International Conference on Multimedia*, ser. MM ’15. New York, NY, USA: ACM, 2015, pp. 5–14.

[2] E. T. Hall, *The hidden dimension*. Garden City, NY: Doubleday, 1910, vol. 609.

[3] K. L. Koay, D. S. Syrdal, M. Ashgari-Oskoei, M. L. Walters, and K. Dautenhahn, “Social roles and baseline proxemic preferences for a domestic service robot,” *International Journal of Social Robotics*, vol. 6, no. 4, pp. 469–488, 2014.

[4] R. Triebel, K. Arras, R. Alami, L. Beyer, S. Breuers, R. Chatila, M. Chetouani, D. Cremers, V. Evers, M. Fiore, *et al.*, "Spencer: A socially aware service robot for passenger guidance and help in busy airports," in *Field and service robotics*. Springer, 2016, pp. 607–622.

[5] F. Yang and C. Peters, "Appgan: Generative adversarial networks for generating robot approach behaviors into small groups of people," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2019, pp. 1–8.

[6] T. Kruse, A. K. Pandey, R. Alami, and A. Kirsch, "Human-aware robot navigation: A survey," *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1726–1743, 2013.

[7] F. Yang and C. Peters, "App-lstm: Data-driven generation of socially acceptable trajectories for approaching small groups of agents," in *Proceedings of the 7th International Conference on Human-Agent Interaction*, 2019, pp. 144–152.

[8] F. Yang, W. Yin, T. Inamura, M. Björkman, and C. Peters, "Group behavior recognition using attention- and graph-based neural networks," in *Proceedings of the 24th European Conference on Artificial Intelligence*, 2020.

[9] L. D. Riek, "Wizard of oz studies in hri: a systematic review and new reporting guidelines," *Journal of Human-Robot Interaction*, vol. 1, no. 1, pp. 119–136, 2012.

[10] F. Yang and C. Peters, "Social-aware navigation in crowds with static and dynamic groups," in *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*. IEEE, 2019, pp. 1–4.

[11] J. Ho and S. Ermon, "Generative adversarial imitation learning," in *Advances in neural information processing systems*, 2016, pp. 4565–4573.

[12] A. Kendon, *Conducting interaction: Patterns of behavior in focused encounters*. CUP Archive, 1990, vol. 7.

[13] X.-T. Truong and T.-D. Ngo, "Dynamic social zone based mobile robot navigation for human comfortable safety in social environments," *International Journal of Social Robotics*, vol. 8, no. 5, pp. 663–684, 2016.

[14] C. Pedica and H. H. Vilhjálmsson, "Study of nine people in a hallway: Some simulation challenges," in *Proceedings of the 18th International Conference on Intelligent Virtual Agents, IVA 2018, Sydney, NSW, Australia, November 05-08, 2018*, 2018, pp. 185–190.

[15] S. B. P. Samarakoon, M. V. J. Muthugala, and A. B. P. Jayasekara, "Replicating natural approaching behavior of humans for improving robot's approach toward two persons during a conversation," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2018, pp. 552–558.

[16] O. A. I. Ramírez, H. Khambhaita, R. Chatila, M. Chetouani, and R. Alami, "Robots learning how and where to approach people," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*. IEEE, 2016, pp. 347–353.

[17] Y. Gao, F. Yang, M. Frisk, D. Hemandez, C. Peters, and G. Castellano, "Learning socially appropriate robot approaching behavior toward groups using deep reinforcement learning," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2019, pp. 1–8.

[18] R. Mead and M. J. Matarić, "Proxemics and performance: Subjective human evaluations of autonomous sociable robot distance and social signal understanding," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 5984–5991.

[19] C. Peters, F. Yang, H. Saikia, C. Li, and G. Skantze, "Towards the use of mixed reality for hri design via virtual robots," in *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, 2018.

[20] H. Kuzuoka, Y. Suzuki, J. Yamashita, and K. Yamazaki, "Reconfiguring spatial formation arrangement by robot body orientation," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2010, pp. 285–292.

[21] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, "How to approach humans? strategies for social robots to initiate interaction," in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, 2009, pp. 109–116.

[22] F. Yang, W. Yin, T. Inamura, M. Björkman, and C. Peters, "Group behavior recognition using attention-and graph-based neural networks," 2020.

[23] M. Lohse, N. van Berkel, E. M. van Dijk, M. P. Joosse, D. E. Karreman, and V. Evers, "The influence of approach speed and functional noise on users' perception of a robot," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 1670–1675.

[24] Z. Henkel, C. L. Bethel, R. R. Murphy, and V. Srinivasan, "Evaluation of proxemic scaling functions for social robotics," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 3, pp. 374–385, 2014.

[25] C.-M. Huang, T. Iio, S. Satake, and T. Kanda, "Modeling and controlling friendliness for an interactive museum robot." in *Robotics: Science and Systems*, 2014, pp. 12–16.

[26] M. L. Walters, K. Dautenhahn, S. N. Woods, and K. L. Koay, "Robotic etiquette: results from user studies involving a fetch and carry task," in *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2007, pp. 317–324.

[27] A. K. Ball, D. C. Rye, D. Silvera-Tawil, and M. Velonaki, "How should a robot approach two people?" *Journal of Human-Robot Interaction*, vol. 6, no. 3, pp. 71–91, 2017.

[28] L. Takayama and C. Pantofaru, "Influences on proxemic behaviors in human-robot interaction," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 5495–5502.

[29] K. Fischer, L. C. Jensen, S.-D. Suvei, and L. Bodenhagen, "Between legibility and contact: The role of gaze in robot approach," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2016, pp. 646–651.

[30] S. Bhagya, P. Samarakoon, M. Viraj, J. Muthugala, A. Buddhika, P. Jayasekara, and M. R. Elara, "An exploratory study on proxemics preferences of humans in accordance with attributes of service robots," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2019, pp. 1–7.

[31] E. Torta, R. H. Cuijpers, and J. F. Juola, "Design of a parametric model of personal space for robotic social navigation," *International Journal of Social Robotics*, vol. 5, no. 3, pp. 357–365, 2013.

[32] D. Carton, A. Turnwald, D. Wollherr, and M. Buss, "Proactively approaching pedestrians with an autonomous mobile robot in urban environments," in *Experimental Robotics*. Springer, 2013.

[33] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE transactions on Systems Science and Cybernetics*, vol. 4, no. 2, pp. 100–107, 1968.

[34] C. Pedica and H. Vilhjálmsson, "Social perception and steering for online avatars," in *International Workshop on Intelligent Virtual Agents*. Springer, 2008, pp. 104–116.

[35] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations," *Journal of computational physics*, vol. 79, no. 1, pp. 12–49, 1988.

[36] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[37] C. D. Kidd, "Sociable robots: The role of presence and task in human-robot interaction," Ph.D. dissertation, Massachusetts Institute of Technology, 2003.

[38] K. Shinozawa, F. Naya, J. Yamato, and K. Kogure, "Differences in effect of robot and screen agent recommendations on human decision-making," *International journal of human-computer studies*, vol. 62, no. 2, pp. 267–279, 2005.

[39] S. Woods, M. Walters, K. L. Koay, and K. Dautenhahn, "Comparing human robot interaction scenarios using live and video based methods: towards a novel methodological approach," in *9th IEEE International Workshop on Advanced Motion Control, 2006*. IEEE, 2006, pp. 750–755.

[40] C. Ennis and C. O'Sullivan, "Perceptually plausible formations for virtual conversers," *Computer Animation and Virtual Worlds*, vol. 23, no. 3-4, pp. 321–329, 2012.

[41] F. Martinez-Gil, M. Lozano, I. García-Fernández, and F. Fernández, "Modeling, evaluation, and scale on artificial pedestrians: a literature review," *ACM Computing Surveys (CSUR)*, vol. 50, no. 5, pp. 1–35, 2017.

[42] F. Yang, J. Shabo, A. Qureshi, and C. Peters, "Do you see groups? the impact of crowd density and viewpoint on the perception of groups," in *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, 2018, pp. 313–318.

[43] B. Okal and K. O. Arras, "Learning socially normative robot navigation behaviors with bayesian inverse reinforcement learning," in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*. IEEE, 2016, pp. 2889–2895.