Effect of Robot Embodiment on Satisfaction With Recommendations in Shopping Malls

Kazuki Sakai, Yutaka Nakamura, Yuichiro Yoshikawa, and Hiroshi Ishiguro

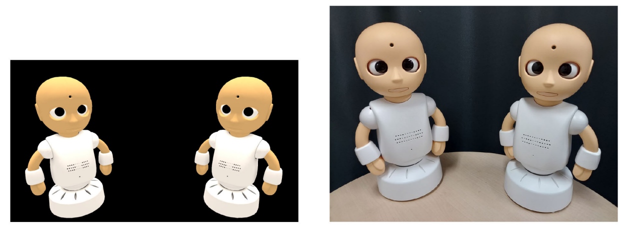
***Abstract*—Recent developments in conversational technologies have attracted researchers to study their applications in recommending items through conversations. It is considered that physical robots, rather than virtual ones, are effective in situations in which robots talk about items near participants. However, in real situations, robots may be required to recommend items that are present but invisible in the scene of the communication. In this study, we conducted a field experiment in a shopping mall to investigate the effects of robot embodiment on recommendation tasks. The robots recommended a dish after talking to participants about their food preferences. We developed a conversational recommendation system and implemented it using physical and virtual robots. The field experiment was conducted in a shopping mall; the visitors were encouraged to participate. The experiment lasted a total of 99 hours (9 hours per day for 11 days) inside and in front of a food court. Although no significant difference in the behavioral aspect was confirmed, the results obtained from 272 conversations suggested that having physical bodies enhanced the satisfaction and agreement with the robots' recommendations.**

***Index Terms*—Social human-robot interaction, acceptability and trust, embodiment, field experiment.**

I. INTRODUCTION

RECENTLY, conversational robots that provide advanced information, such as recommendations, have been developed [1]–[3]. These robots can provide sophisticated services while understanding the user's feelings and background. Conversational robots are expected to be used in daily life situations, for example, as artificial shop clerks [4] and in hospitals [5].

The effectiveness of robot embodiment has been widely studied by researchers in the field of human-robot interaction. Recent surveys have summarized studies concerning the differences between physical and virtual robots [6], [7]. According to these surveys, most studies reported that physical robots are more attractive, persuasive, and positive than virtual ones, whereas some studies showed the opposite results in certain situations. One study showed that physical robots are effective for tasks performed in real situations, whereas virtual robots are effective for tasks performed in virtual situations [8]. According to these results, physical robots are considered to be better than virtual ones for recommending items that physically exist in the place of the communication. However, in real-world applications of robot recommendations, robots are sometimes required to recommend items that are present but invisible in the scene of the communication. For example, when a robot recommends some items at the entrance of a shop, these are not visible to the user or the robot. Therefore, we should investigate whether having a physical body is effective even if the recommended items are not visible to the participants.



(a) (b)

Fig. 1. Humanoid robots: (a) virtual robots and (b) physical robots.

In this study, we conducted a field experiment in a shopping mall to investigate the effects of robot embodiment recommendation tasks. The robots recommend a dish after talking to a participant about their food preferences. We developed a multi-robot dialogue system that could talk to a participant while successively estimating the next item to be questioned or recommended based on the disclosed preferences in previous questions. In the experiment, we implemented the developed dialogue system in both virtual (Fig. 1(a)) and physical robots (Fig. 1(b)). Visitors to the shopping mall were encouraged to attend the experiment. We evaluated their satisfaction with the interaction by analyzing their questionnaire responses and their agreement with robot recommendations by analyzing their utterances. Moreover, we examined whether the participants ate or planned to eat the recommended dish after the conversation with the physical or virtual robots.

The remainder of this paper is organized as follows. In Section II, related work investigating the advantages and disadvantages of having a physical body is described. Section III

Manuscript received July 14, 2021; accepted October 22, 2021. Date of publication November 16, 2021; date of current version November 24, 2021. This work was supported in part by JST ERATO under Grant JPMJER1401, in part by JSPS KAKENHI under Grants JP19H05691 and JP20K23344, and in part by the Innovation Platform for Society 5.0 (JPMXP0518071489), Japan. *(Corresponding author: Kazuki Sakai.)*

Kazuki Sakai, Yuichiro Yoshikawa, and Hiroshi Ishiguro are with the Graduate School of Engineering Science, Osaka University, Toyonaka 560-8531, Japan (e-mail: sakai.kazuki@irl.sys.es.osaka-u.ac.jp; yoshikawa@irl.sys.es.osaka-u.ac.jp; ishiguro@irl.sys.es.osaka-u.ac.jp).

Yutaka Nakamura is with the Guardian Robot Project, RIKEN Information R&D and Strategy Headquarters, RIKEN, Kyoto 619-0288, Japan (e-mail: yutaka.nakamura@riken.jp).

Digital Object Identifier 10.1109/LRA.2021.3128233

describes the development of the dialogue system for recommendation. In Section IV, we describe the method and results of the field experiment in a shopping mall. In Section V, the results are discussed, and the conclusions are presented in Section VI.

II. Related Work

Many studies have focused on the embodiment of a robot. Such studies can be divided into two types: studies on whether the appearance should be humanlike and on whether the body should be physical. In the former type of studies, humanlike robots were compared to smart speakers [9] or text chatbots [10]. In the current study, we focused on the latter type: a comparison between physical and virtual robots. Many studies have shown the positive effects of robots' physical bodies on task-oriented interactions (see review [6], [7]). However, because all studies in these reviews were limited to laboratory experiments, the effectiveness in real-world applications is unclear. Nishio *et al.* conducted a laboratory experiment that was similar to a field experiment. They found that physical robots were more effective than virtual ones in communication when they succeeded in speech recognition. However, this study focused only on the elderly [11].

In the context of the robot's recommendation and persuasion, studies have investigated the effect of embodiment. Schneider *et al.* showed that, in a robot application to encourage a user to exercise, the time users spent exercising was longer with the physical robot than with the virtual one, and the former was preferred over the latter consistently [12]. Wang *et al.* showed that for a decision-making task, a physical robot could influence the user's decision more than an onscreen agent with an identical appearance [13]. However, this effect of physical embodiment is not consistent across situations. Thellman *et al.* reported that the social presence of an agent was a more influential factor for user's acceptance of it in an ultimatum game rather than the agent's embodiment [14]. An online survey to compare the performance of food recommendation by a tablet device, a humanoid robot, and a human suggested that the acceptance rate was improved only in the human condition, but the effect depended on the specific recommendation scenario [15]. Therefore, the effectiveness of the robot's embodiment is unclear in our target situations where the robots recommended items that were not displayed in front of the participants.

III. Dialogue System

We developed a dialogue system with multiple robots that recommended dishes while asking the participant's preferences. We adapted two robots into the dialogue system because multiple robots improved the participants' tolerance to continuing the conversation even if recognition errors occurred [16], [17]. To select the next items to ask about and recommend, the robots successively estimated the participants' preferences by using Gaussian process regression [18] based on a similarity model.

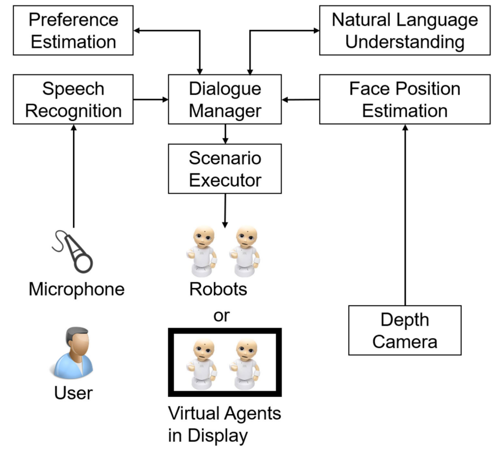


Fig. 2. System architecture.

A. Preference Estimation

To estimate the user's preference, we used subjective similarity, which is a relational model of items reflecting a certain person's criteria. Subjective similarity is useful for calculating the similarity between items of different categories, such as foods and feelings. Subjective similarity is also useful for robots to estimate preferences in situations where they do not have the user's information before the conversation.

Subjective similarity was created using the following procedure. First, we collected the similarity data of a single person by using the crowd clustering method [19], which divides the entire task into small tasks to facilitate data input. Then, the data obtained by the tasks are integrated and analyzed by an infinite relational model [20], which obtains their similarity relationships, that is, subjective similarity.

To estimate preferences based on subjective similarity, we used a Gaussian process regression [18], which obtains the estimated preference values of the items. Based on these values, we chose the next items to ask about and recommend. Here, we focused on the history of the values; that is, the item whose estimated value was most changed from the previous estimated value was chosen. When this strategy is used, the participant can recognize that the current item the robot asked about is related in some way to the previous items.

B. System Architecture

Fig. 2 shows the system architecture of the dialogue system. The user utterances are recognized by the speech recognition module, and the result is sent to the dialogue manager. The speech recognition works only when the robots wait for the answer. The dialogue manager obtains the information from preference estimation and natural language understanding modules. Then, it generates the dialogue scenario consisting of multimodal components: the robots' utterances, gaze targets, and gestures. That is, given the components, it generates dialogue scenarios consisting of a command sequence for the robots with the information of when it should be sent to the executor. Note that some utterances were prepared as a template with

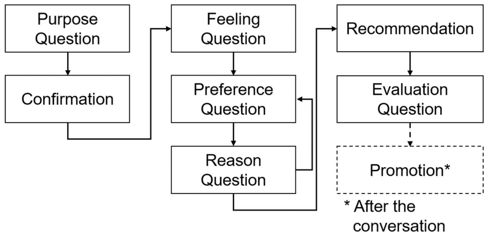


Fig. 3. Dialogue flow.

variable word slots and completed with adequate words, such as item names. To avoid giving a monotonous impression, some utterances involve variant expressions to be randomly chosen. The scenario executor schedules the commands and sends them to the robots as scheduled.

The natural language understanding module performs morphological analysis of utterance texts and then recognizes keywords, including agreement/disagreement, and comments. Keyword recognition is performed based on the word match method. In advance, we manually created a list of all possible words that the user could say to represent agreement or disagreement, such as “yes” and “like” for agreement and “no” and “dislike” for disagreement. If the results of the morphological analysis include such prepared words, the module outputs the represented category (i.e., agreement/disagreement). For comment recognition, the module extracts the content words (noun, adjective, and verb) from the results of the morphological analysis and repeats the words. The preference estimation module estimates the preferences of items based on the preference estimation method described Section III-A and chooses the next item to be asked about by the robots. To establish eye contact between the robots and user, the face position estimation module estimates the position of the user’s face by using a depth camera and updates the gaze target.

Fig. 3 shows the dialogue flow. First the robots asked a question about the purpose of the participant’s visit. Then the robots asked for confirmation to continue the dialogue. If the participant disagreed with continuing the dialogue, the dialogue was ended. Then the robots asked a question about the participant’s current feeling. After that, the robots asked questions about participant’s food preferences and reasons for them. After repeating the preference questions twice, the robots recommended a dish in consideration of the obtained answers. After that, the robots asked a question about recommendation evaluation. In the closing part of the dialogue, the robots suggested eating the recommended dish in the food court and asked the participant to complete the questionnaire. To encourage the participants to plan to eat the recommended meal, the robots talked to each other about their impressions of the dish after the conversation. In all questions, when the robots could not recognize the voice of the participant because of the noise or long pause, the robots repeated the same question. Fig. 4 shows an example of the dialogue.

Speaker	Utterance
R1	Hello, what is your purpose in Expocity today?
H	Buying clothes.
R1	I see, there are many shops.
R2	What do you plan shopping for?
H	No plan.
R2	I see.
R1	So, may I recommend a dish?
H	OK.
R1	Thank you! First, are you hungry now?
H	Yes.
R2	So do you like Yakisoba?
H	Yes, I do.
R1	Oh, why do you like it?
H	I like it because the Japanese brown sauce is good.
R1	Japanese brown sauce is good, I see.
R2	So, I think you definitely like carbonara, don’t you?
H	Yes.
R1	I expected that. Why do you like it?
H	The creamy sauce.
R1	The creamy sauce, I see.
R2	I know you are hungry...
R1	We picked a dish to recommend. Would you eat special Akamaru (a type of ramen)?
H	I will try it.
R1	Please try it!
R2	You can get it at that place called Ippudo.
R1	How would you rate our recommendation out of 100?
H	I think 90 points.
R1	90 points, I’m glad to get such a high score.
R2	Thank you for talking with us. If you eat lunch, please fill in the questionnaire after eating.
R1	If not, please fill it in now.
R2	Please come again! Good-bye.
R1 to R2	I’m glad to have our recommendation accepted. Special Akamaru is good because of the thick soup.
R2 to R1	I agree. In addition, chewy noodles are also good.

Fig. 4. Dialogue example. Speaker R1, R2, and H refer to Robot-1, Robot-2, and Human, respectively.

IV. Experiment

To verify the effects of robot embodiment on recommendation acceptance in a shopping mall, we conducted a field experiment at two locations, inside and in front of a food court of LaLaport Expocity (a shopping mall in Osaka, Japan) from 11 am to 8 pm on 15–25 March, 2020. All procedures in this experiment were approved by the ethical committee of the Graduate School of Engineering Science, Osaka University, Japan. We compared two conditions: Physical and Virtual. One location was assigned Physical condition and the other Virtual condition. The conditions were switched daily. We evaluated the satisfaction and agreement with the recommendations. In addition, we examined whether the participants ate or planned to eat the recommended meal after a conversation with physical or virtual robots.

A. Method

1) Participants: A total of 440 visitors to the shopping mall (234 people in the Physical condition, and 186 people in the Virtual condition) participated in the experiment. We only used data obtained from subjects who were over 13 years old and participated in the experiment for the first time. As a result, the data of 137 participants (36 men, 78 women, and 23 undisclosed) from Physical condition and 135 participants (47 men,

(a) Physical condition (b) Virtual condition

Fig. 5. Experimental setup: (a) physical condition and (b) virtual condition.

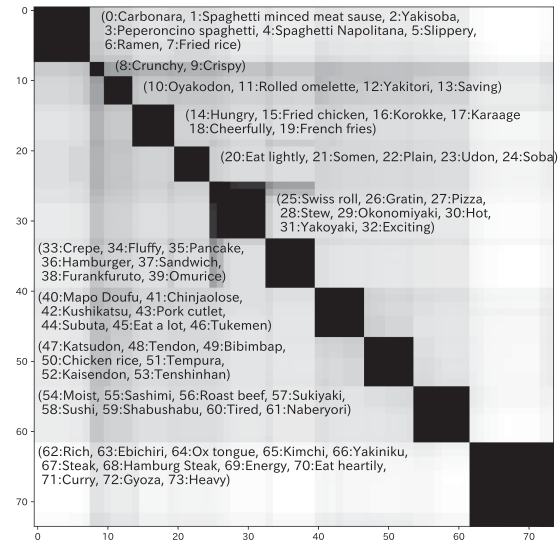


Fig. 6. Subjective similarity matrix. The rows and columns of the matrix are the IDs of the items. The shade of the matrix represents the degree of similarity between the items: black indicates high similarity, and white indicates dissimilarity. Note that the similarity is a continuous number ranging from zero to one.

71 women, and 17 undisclosed) from Virtual condition were used.

2) Appratus: Fig. 5 shows the setups of the event booths. Three desks were placed in each booth. A laptop PC with touch display for acquiring the informed consent and providing the instructions, a printer, and 2D code reader were placed on the first desk. Two robots or a display, another laptop PC, a microphone, a depth camera (Intel RealSense D435i), a pair of stereo speakers, and a 2D code reader were placed on the second desk. A box to collect the questionnaires and gifts for participation in the experiment were placed on the third desk. In Physical condition, two table-top humanoid robots called CommU (developed by the collaboration between Osaka University, Japan, and VStone Co., Ltd.) were used. In Virtual condition, we used a display (FlexScan EV2750 27 in, EIZO) with a screen sufficient to draw two virtual agents with the same shape and size as CommU. To balance the positional relationship of the robots from the viewpoint of the subjects, we adjusted the scene camera (viewpoint of 3D rendering) at the top in front of the robots. To distinguish the agents' utterances, their utterances were divided between the stereo speakers based on their position. To emphasize to the participants that all robots/agents were different, a name plate was set on the robots' chests in both conditions. When the robots spoke, their mouth opened and closed in synchrony with the voice. With 14 degrees of freedom, they could produce various nonverbal gestures such as nodding.

3) Stimuli: The dialogue flow was the same in both conditions. The duration of the dialogue was approximately five minutes. Note that while there were no participants, the robots talked to each other about food, experimental setup, and daily life every minute.

In this study, we adapted the dishes of the restaurants in the food court for the recommendations. Two dishes from 14 places (28 dishes in total), from which we obtained a permission to use their information, were chosen as the recommendation items. To create subjective similarity, all dishes were labeled as food categories. As a result, 21 labels were used. In addition, 18 expressions for the feelings related to eating a meal and an additional 35 food items were used. For the final 74 items, a single person in our laboratory repeated 200 small tasks to input the similarity data. The subjective similarity used in the experiment is shown in Fig. 6.

To link the questionnaire and system logs, we used a 2D code. When the participants agreed to participate in the experiment on instruction graphical user interface (instruction GUI), the GUI generated a unique ID. The ID, condition, and timestamp were encoded into a 2D code, and a questionnaire paper with that code was printed. Using the 2D code, the system could link the ID and the conversation.

4) Procedure: First, the participants read the instructions on the instruction GUI and decided whether to participate in the experiment. The instructions described the purpose of the experiment, precautions, flow of the experiment including the questionnaire, the way to leave the experiment, and the treatment of the acquired data. If they decided to participate in the experiment, they pressed "participation" button placed on the last page of the instruction GUI. If not, they pressed "decline" button and left here. The subjects then received the printed questionnaire with the 2D code and moved to the next desk where the robots/display was placed. The participants then read instructions that contained guidance on talking to the robots. After reading the instructions, the participants held the 2D code over the 2D code reader, which triggered the start of the conversation. At the end of the conversation, the robots requested participants to complete questionnaires using one of two options (see Fig. 4). If the participants planned to go to the food court, they brought the printed questionnaire with them. After finishing eating, they completed the questionnaire and returned to the booth to place it in the box. If the participants did not have plans to eat, they completed the questionnaire immediately and submitted it.

5) Evaluation: The same questionnaire was used for all participants. The questionnaire consisted of two parts: two items about gender and age and seven items about the participants' impressions of the dialogue. Two of the seven items queried the experience with the experiment.

Q1 Did you talk to these robots?

Q2 Did you talk to the robots in the other location?

For Q1, the participants selected from three options: one, two, and more time(s). For Q2, they selected from four options: no, one, two, and more time(s).

The next two items queried the eating activity:

Q3 Did you eat something in the food court?
Q4 Did you eat the recommended dish today?

We asked Q3 to check if the participants had eaten a meal before the conversation with the robots. That is, we asked if (i) they had eaten before the conversation, (ii) they ate after the conversation, or (iii) they have not eaten yet. We asked Q4 to check if the participants accepted the robots' recommendation. That is, we asked if (i) they have eaten the recommended food, (ii) planned to eat the recommended food, or (iii) have not eaten the recommended food yet. We supposed that persons who have not eaten it yet but planned to do so later chose the second option, while those who have not eaten it and did not plan to do so later chose the third option.

The remaining three items queried the satisfaction with the conversation:

Q5 Did you enjoy the conversation with the robots?
Q6 Did you feel that the robot's recommendation was persuasive?
Q7 Do you wish to talk to the robots again?

For Q5-7, they rated the items on a seven-point Likert scale ranging from 1 ("strongly disagree") to 7 ("strongly agree"). The midpoint value of four corresponded to "undecided". We calculated the average scores of the satisfaction items.

To evaluate the agreement with the recommendations, we used the score that was provided as an answer to the last question of the conversation, that is, "How would you rate our recommendation out of 100?". Note that we used only scores that were successfully recognized by speech recognition.

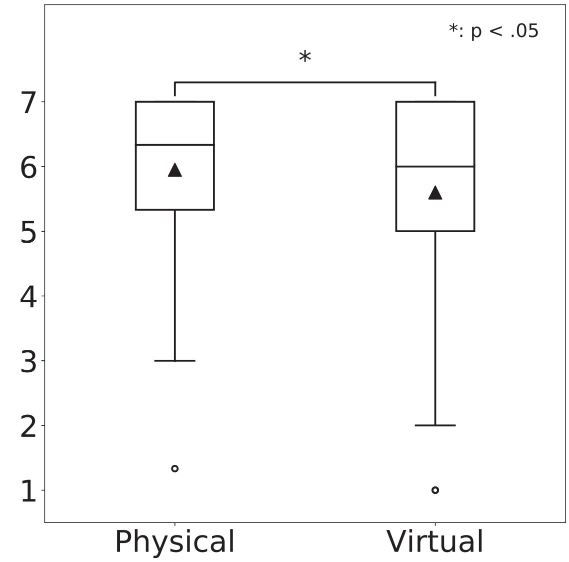


Fig. 7. Boxplots of the average score of satisfaction.

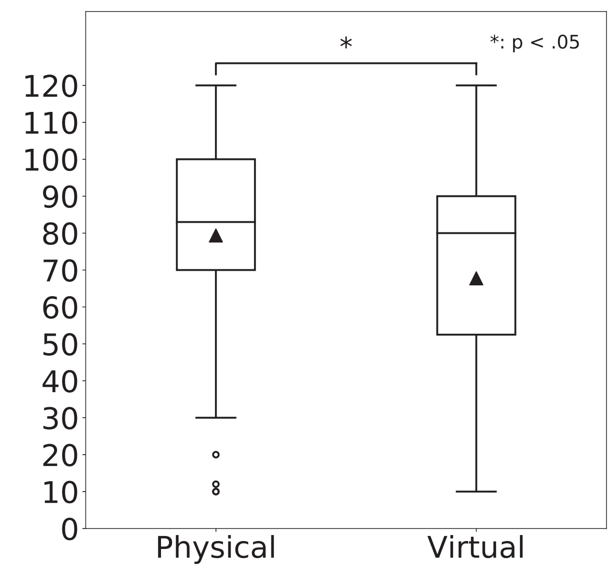


Fig. 8. Boxplots of the scores of agreement with the recommendation.

B. Results

First, we checked the smoothness of the dialogue. We compared the number of times that the questions were repeated when the dialogue system did not obtain any utterances from the speech recognition using a t-test. As a result, we found no significant difference between conditions (Physical condition: $M = 2.190$, $SD = 2.315$; Virtual condition: $M = 1.817$, $SD = 3.235$; $t(124) = 0.813$, $p = 0.418$).

Fig. 7 shows the boxplots of the average scores of satisfaction with the dialogue. Note that Cronbach's alpha of the three items about satisfaction was 0.91. The average score of satisfaction in Physical condition was found to be significantly higher than that in Virtual condition (Physical condition: $N = 130$, $M = 5.951$, $SD = 1.106$; Virtual condition: $N = 131$, $M = 5.599$, $SD = 1.413$; $t(245) = 2.242$, $p = 0.026$).

For Q3, 24 participants in Physical condition and 15 participants in Virtual condition answered, "after conversation". Among these participants, the number of people that answered "have eaten" or "plan to eat" for Q4 was 6 and 5, respectively. We compared the ratio (6/24 vs. 5/15) using a two-sample proportion test. However, we did not find a significant difference between the conditions ($\chi^2(1) = 0.317$, $p = 0.574$). In contrast, in Q3, 32 and 36 participants answered that they had eaten "before the conversation" in Physical and Virtual conditions, respectively. Moreover, 80 and 82 participants answered that they "have not eaten yet" in Physical and Virtual conditions, respectively. In the sample of participants who have not eaten, there was no significant difference in the ratio of persons choosing the option of "plan to eat" in Q4 between Physical (3 out of 80) and Virtual (7 out of 82) conditions ($\chi^2(1) = 1.602$, $p = 0.206$). In the sample of users who had eaten before the conversation, there was no significant difference in the same ratio between Physical (3 out of 32) and Virtual (2 out of 36) conditions ($\chi^2(1) = 0.363$, $p = 0.547$).

Fig. 8 shows the boxplots of the scores of the participants' agreement with the recommendation, which is successfully recognized by the speech recognition. The score of Physical condition was found to be significantly higher than that of Virtual condition (Physical condition: $N = 53$, $M = 79.358$, $SD = 26.328$; Virtual condition: $N = 50$, $M = 67.760$, $SD = 31.230$; $t(96) = 2.032$, $p = 0.045$).

V. Discussion

No significant difference in the number of repeated questions indicated that the participants' impressions of the robots' capabilities for understanding the participants' utterances were not biased between the conditions. Analyses of questionnaire scores revealed that the participants' satisfaction and agreement with the robot's recommendation was enhanced by the robots having physical bodies. The current result suggests that the positive effect of the physical embodiment reported in a previous laboratory experiment [8], [21] remains valid even in a real-world recommendation situation, where the recommended item is not displayed in the location of the conversation. This will accelerate the use of physical robots in real situations, although the problem of high cost remains. However, it is worth noting that the result regarding the agreement of the recommendation is limited because the agreement score was obtained through speech recognition; therefore, we considered only the partial cases in which the participant's answer was successfully identified. In addition, it should be clarified whether eating the recommended food, eating other food, or not eating anything before talking to the robots or completing the questionnaire influences the results of the questionnaire and the agreement with the recommendation. For example, if the participants eat the recommended food, they might provide a higher evaluation rating. Conversely, if they eat food that the robots did not recommend, they might give a lower evaluation score. However, owing to the small sample size, it is difficult to analyze this effect. To examine it, we would need to encourage the participants to complete the questionnaire after having a meal by providing an incentive, such as a discount. Moreover, we would require a mechanism to monitor which dish the participant ate.

However, we did not find a significant difference in the ratio of participants who ate or planned to eat the recommended dish after the conversation. A possible reason is the time of the experiment, which did not involve only the lunch and dinner time but also the time between; however, the food court provided menus suitable for lunch and dinner. Therefore, in the in-between periods, it was difficult for the robots to motivate participants to eat the recommended meal even if the recommendation was successfully provided in terms of the participants' impressions. Another possible reason for the non-significant difference in this measure is the choice of topic to be recommended. A previous study [22] revealed that users did not prefer to talk about non-robot-friendly topics, such as food and art, with conversational robots. This might decrease the influence of the robot recommendation in the current experiment, where the robots recommended a food dish to be eaten. In the future, it is worth examining the recommendation effects with more robot-friendly topics, such as books and electrical appliances, which are understandable by the sensing modalities available to the current robots.

The number of participants who planned to eat something and accepted the recommended dish (6/24 and 5/15) was small. We consider that the low numbers happened because of the bad choice of time for experiment including the period between lunch and dinner as well as the potential non-robot-friendliness of the topic of food. On the other hand, although these success rates (25% and 33% in Physical and Virtual conditions, respectively) were not large and should be improved in the future work, we did not consider that they are quite low or hopeless because the chance level was 7% in this study, which is the probability of randomly selecting one restaurant from the 14 participating ones. If the results were from a reliable analysis with sufficient number of samples, it could imply that the robots' recommendations have influenced the participants' food selection. To confirm this, we need to increase the number of conversations with the participants who had not eaten food before the conversation. However, in the current implementation, the robots did not actively encourage the visitors to participate in this experiment. In the future, we will develop new technology for attracting visitors. In addition, to evaluate the success rate of the current dialogue system, it is worth conducting further experiments to compare it to humans or other media.

This study had some limitations. In this study, we used virtual robots on the screen to be compared to physical ones; however, a physical robot displayed on the screen could be another candidate for comparison. Li *et al.* stated that the essential differences between physical and virtual robots are whether they have physical bodies and whether their bodies are shown on the monitor [6]. In the future, we should also include the condition of a physical robot displayed on the screen to investigate how these two factors contribute to the impressions. Second, we used only one subjective similarity model created based on one person's criteria. Therefore, the participants may consider the order of the questions strange. It is worth developing a mechanism to adapt the similarity model and examine its effect on the recommendation conversation. Third, we evaluated the behavioral metric with a questionnaire. To optimize the current dialogue systems as the recommendation system, we need to define a reliable behavioral metric to check if the participants purchased the recommended items. Therefore, it is worth conducting future experiments with the extended dialogue systems not only recommending items but also directly selling them.

VI. Conclusion

We investigated the embodiment of robots in a shopping mall. We conducted a field experiment in which physical or virtual robots recommended the dishes to visitors of the food court. Our results suggest that physical embodiment enhances the participant satisfaction and agreement with recommendations. However, we did not find a significant difference in behavior change after the conversation with the robots. The results will encourage not only researchers but also businesspeople to introduce physical robots into daily life situations, even though the cost problem remains. In future work, we will investigate the influence of embodiment on behavior change and of another type of agent, such as a physical robot displayed on a screen.

Acknowledgment

We would like to thank LaLaport EXPOCITY for use of facilities.

References

[1] Z. Yang, G. Levow, and H. Meng, "Predicting user satisfaction in spoken dialog system evaluation with collaborative filtering," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 8, pp. 971–981, Dec. 2012.

[2] L. Martínezt, L. G. Pérez, M. Barranco, and M. Espinilla, "Improving the effectiveness of knowledge based recommender systems using incomplete linguistic preference relations," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 16, no. supp02, pp. 33–56, 2008.

[3] W.-N. Zhang, Q. Zhu, Y. Wang, Y. Zhao, and T. Liu, "Neural personalized response generation as domain adaptation," *World Wide Web*, vol. 22, no. 4, pp. 1427–1446, Jul. 2019.

[4] M. Watanabe, K. Ogawa, and H. Ishiguro, "Can androids be salespeople in the real world?" in *Proc. 33rd Annu. ACM Conf. Extended Abstr. Hum. Factors Comput. Syst.*, 2015, pp. 781–788.

[5] E. Fosch-Villaronga, H. Felzmann, M. Ramos-Montero, and T. Mahler, "Cloud services for robotic nurses? assessing legal and ethical issues in the use of cloud services for healthcare robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 290–296.

[6] J. Li, "The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents," *Int. J. Hum.- Comput. Stud.*, vol. 77, pp. 23–37, 2015.

[7] E. Deng, B. Mutlu, and M. Mataric, "Embodiment in socially interactive robots," *Found. Trends Robot.*, vol. 7, pp. 251–356, 2019.

[8] K. Shinozawa, F. Naya, J. Yamato, and K. Kogure, "Differences in effect of robot and screen agent recommendations on human decision-making," *Int. J. Hum.- Comput. Stud.*, vol. 62, no. 2, pp. 267–279, 2005.

[9] D. Kontogiorgos, S. van Waveren, O. Wallberg, A. Pereira, I. Leite, and J. Gustafson, "Embodiment effects in interactions with failing robots," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2020, pp. 1–14.

[10] F. Correia, S. Gomes, S. Mascarenhas, F. S. Melo, and A. Paiva, "The dark side of embodiment - teaming up with robots vs disembodied agents," in *Proc. Robot.: Sci. Syst.*, Corvalis, OR, USA, Jul. 2020, doi: 10.15607/RSS.2020.XVI.010.

[11] T. Nishio *et al.*, "The effects of physically embodied multiple conversation robots on the elderly," *Front. Robot. AI*, vol. 8, p. 61, 2021.

[12] S. Schneider and F. Kummert, "Comparing the effects of social robots and virtual agents on exercising motivation," in *Proc. Int. Conf. Social Robot.*, 2018, pp. 451–461.

[13] B. Wang and P.-L. P. Rau, "Influence of embodiment and substrate of social robots on users' decision-making and attitude," *Int. J. Social Robot.*, vol. 11, no. 3, pp. 411–421, 2019.

[14] S. Thellman, A. Silvervarg, A. Gulz, and T. Ziemke, "Physical vs. virtual agent embodiment and effects on social interaction," in *Proc. Int. Conf. Intell. Virtual Agents*, 2016, pp. 412–415.

[15] S. Herse *et al.*, "Bon appetit! robot persuasion for food recommendation," in *Proc. Companion ACM/IEEE Int. Conf. Hum.-Robot Interact.*, 2018, pp. 125–126.

[16] T. Arimoto, Y. Yoshikawa, and H. Ishiguro, "Multiple-robot conversational patterns for concealing incoherent responses," *Int. J. Social Robot.*, vol. 10, no. 5, pp. 583–93, 2018.

[17] H. Sugiyama, T. Meguro, Y. Yoshikawa, and J. Yamato, "Avoiding breakdown of conversational dialogue through inter-robot coordination," in *Proc. 17th Int. Conf. Auton. Agents MultiAgent Syst.*, 2018, pp. 2256–2258.

[18] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: Univ. Press Group Limited, 2006.

[19] R. Gomes, P. Welinder, A. Krause, and P. Perona, "Crowdclustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, no. 24, pp. 558–566.

[20] C. Kemp, J. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda, "Learning systems of concepts with an infinite relational model," in *Proc. 21st Nat. Conf. Artif. Intell.*, 2016, pp. 381–388.

[21] J. Fasola and M. J. Matarić, "A socially assistive robot exercise coach for the elderly," *J. Hum.-Robot. Interact.*, vol. 2, no. 2, pp. 3–32, 2013.

[22] T. Uchida, T. Minato, and H. Ishiguro, "The relationship between dialogue motivation and attribution of subjective opinions to conversational androids," *Trans. Japanese Soc. Artif. Intell.*, vol. 34, no. 1, pp. 1–8, 2019. [Online]. Available: https://www.jstage.jst.go.jp/article/tjsai/34/1/34_B-I62/_article/-char/en