Investigation of the impression given by the appearance and gestures of a virtual reality agent describing a display product

Michiko Inoue[1], Shouta Hioki[1], Fuyuko Iwasaki[1], Shunsuke Yoneda[1], and Masashi Nishiyama[1][0000−0002−5964−3209]

Graduate School of Engineering, Tottori University
101 Minami 4-chome, Koyama-cho, Tottori, 680-8550, Japan
mi.inoue@tottori-u.ac.jp

Abstract. In recent times, there has been a rise in virtual reality agents employed during virtual product exhibitions to describe a display product to users. Previous analytical studies investigated how the agent's appearance influences a user's impression in virtual product exhibitions. Here, we consider how the agent's gestures in addition to their appearance influences a user's impression. In this paper, we investigate the impression that the user perceives from the appearance and gestures of a virtual reality agent describing a display product to the user. Experimental results revealed that users' impressions were influenced by both the agent's appearance and gestures.

Keywords: VR Agent · Gestures · Appearance · Subjective Assessment · Impression.

1 Introduction

Recently, the use of virtual reality (VR) agents in virtual product exhibitions has emerged. The VR agent introduces a product on display to a user in order to sell it. It is known that the VR agent's description influences a user's willingness to purchase a product [3]. However, when the VR agent is describing the product to the user, the user's impression of the VR agent is also important. For example, a user is more likely to purchase a product if the user has a good impression of the VR agent. Therefore, this paper discusses the impression of the VR agent that the user perceives when the VR agent describes a product to the user.

We aim to understand how the VR agent describes the product display influences the impression perceived by the users. There are many factors that affect this impression, such as the VR agent's appearance, gestures, conversation content, and voice quality. In this paper, we focus on the VR agent's appearance and gestures, as shown in Fig. 1, which strongly affect the impression of the VR agent through human visual perception. Existing analytical studies [2, 4] reported that the VR agent's appearance affects the user's willingness to purchase when the VR agent is describing a display product. However, these existing analytical

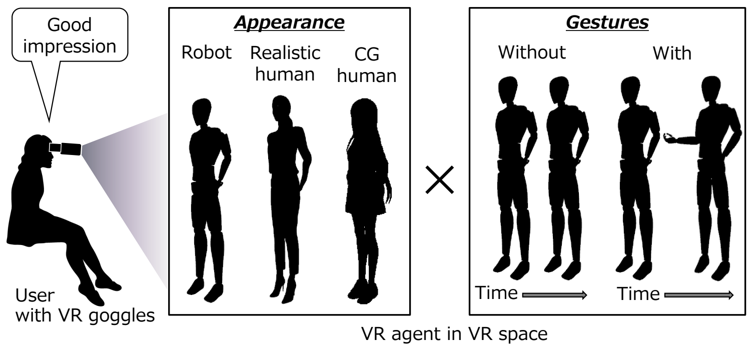

Fig. 1. Example of an agent bowing and pointing with their hand during a product description.

studies did not consider how the combination of the VR agent's appearance and gestures affects the user's impression.

In this paper, we investigated the following hypotheses regarding the impression that the user perceives from the appearance and gestures of a VR agent describing a display product to the user.

H_1 : The VR agent's appearance changes the strength of the good impression that the user perceives.
H_2 : The VR agent's gestures change the strength of the good impression that the user perceives.
H_3 : Combining the VR agent's appearance and gestures changes the strength of the good impression that the user perceives.

In the following, we explain the experimental design of the agent's appearance and gestures in VR space in Section 2, present the results of the subjective evaluation of the experimental subjects in Section 3, and summarize the study in Section 4.

2 Experimental design

2.1 Overview

In this paper, we employ a two-factor analysis of variance to validate the three hypotheses: H_1, H_2, and H_3. In the analysis, if we observe the main effect of the agent's appearance, we support H_1; if we observe the main effect of the agent's gestures, we support H_2; and if we observe an interaction between the agent's appearance and the agent's gestures, we support H_3. We refer to users as "subjects" in the following. We also refer to an object described by the agent in virtual product exhibitions as a "display product."

We investigated three hypotheses concerning the description of a display product within a VR environment, exploring interactions between a VR agent

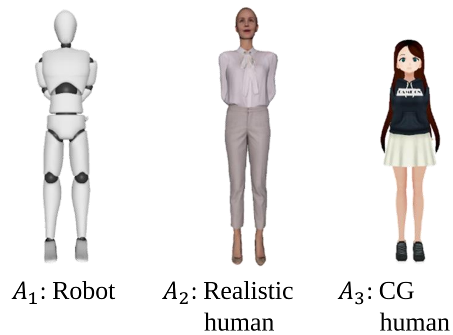

Fig. 2. Example of the three type of agents used for product description.

and subjects. Specifically, we emulate interactions in which a human clerk agent describes a display product to subjects in the physical space. This study focuses on describing a display product in one-on-one interactions, which represent the minimum number of subjects. When describing a display product with two or more agents and subjects, it is essential to consider the impact of the relationships between the agents and subjects. Therefore, we initially experimented under conditions that minimize such influences.

2.2 Agent appearance

We consider the VR agent's appearance when investigating H_1, H_2, and H_3. A VR agent's appearance can exhibit considerable variation, ranging from humanoid to animal-like because of the extensive creative possibilities allowed within the VR environment. In this study, we opted for humanoid agents with a human skeletal structure to reproduce the gestures performed by human agents in the real world. While a diverse array of humanoid agents are used in virtual product exhibitions, we categorized them into three primary agent types.

A_1**:** Robot
A_2**:** Realistic human
A_3**:** Computer graphics (CG) human

The reasons for using the realistic human agent include the potential for subjects to feel as if they are receiving descriptions from a real human. The reasons for using the CG human agent include the lightweight nature of 3D data and the ease with which CG human model designers can convey their intentions to subjects; for example, by giving the VR agent a cute appearance. A robot agent is used because an agent possessing a human skeleton but lacking a face and clothing may yield effects distinct from typical human interactions. This study substantiates the hypotheses through the evaluation of these three agent types. Figure 2 presents examples of the three agents evaluated in this experiment.

2.3 Agent gestures

We consider the gestures of the VR agent. Generally, the procedure an agent uses to explain a display product involves approaching the subject, greeting them, and then explaining the characteristics of a display product. In this study, we focused on display product descriptions in which the agent and the subjects are likely to engage in conversation rather than one in which the agent simply moves around. Therefore, we refer to the procedure of greeting the subjects and explaining a display product as "describing a display product." In the experiments, we incorporated gestures commonly used when describing a display product in the real world.

When describing a display product, an agent can perform gestures indicating the product, the subject, or the agent itself. Examples of an agent's gestures include bowing to express respect for the subject or identify the recipient of the display product description, pointing in the direction of the product, and making gestures suggesting happiness or success to convey the agent's feelings to the subjects. In this study, we considered gestures towards the products and the subjects, and excluded gestures aimed at the agent itself. During the validation of hypotheses H_1, H_2, and H_3, the agent describes the subjects using both a bowing gesture towards the subject and a pointing gesture towards the products. The agent's gesture conditions are as follows:

$G_{w/o}$**:** Without gestures
$G_{w/}$**:** With gestures

2.4 Stimuli

We designed six comparison conditions to validate the hypotheses by combining the agent's appearance, as described in Section 2.2, and the agent's gestures, as described in Section 2.3. These comparison conditions are as follows:

M_1**:** Robot without gestures (A_1, $G_{w/o}$)
M_2**:** Robot with gestures (A_1, $G_{w/}$)
M_3**:** Human without gestures (A_2, $G_{w/o}$)
M_4**:** Human with gestures (A_2, $G_{w/}$)
M_5**:** CG human without gestures (A_3, $G_{w/o}$)
M_6**:** CG human with gestures (A_3, $G_{w/}$)

We generated stimuli for all comparison conditions. Figure 3 shows examples of a stimulus for each condition. We controlled the timing and manner of the gestures across comparison conditions M_2, M_4, and M_6.

While describing a display product, the agent's gestures typically occur concurrently with the conversation. In this study, in addition to generating the agent's gestures, we incorporated verbal explanations. The agent's verbal content included the following phrases: "Welcome," "Today's recommendation is the two-seater sofa," and "Please consider buying it." The verbal content was presented in the above order. During the description of a display product, the

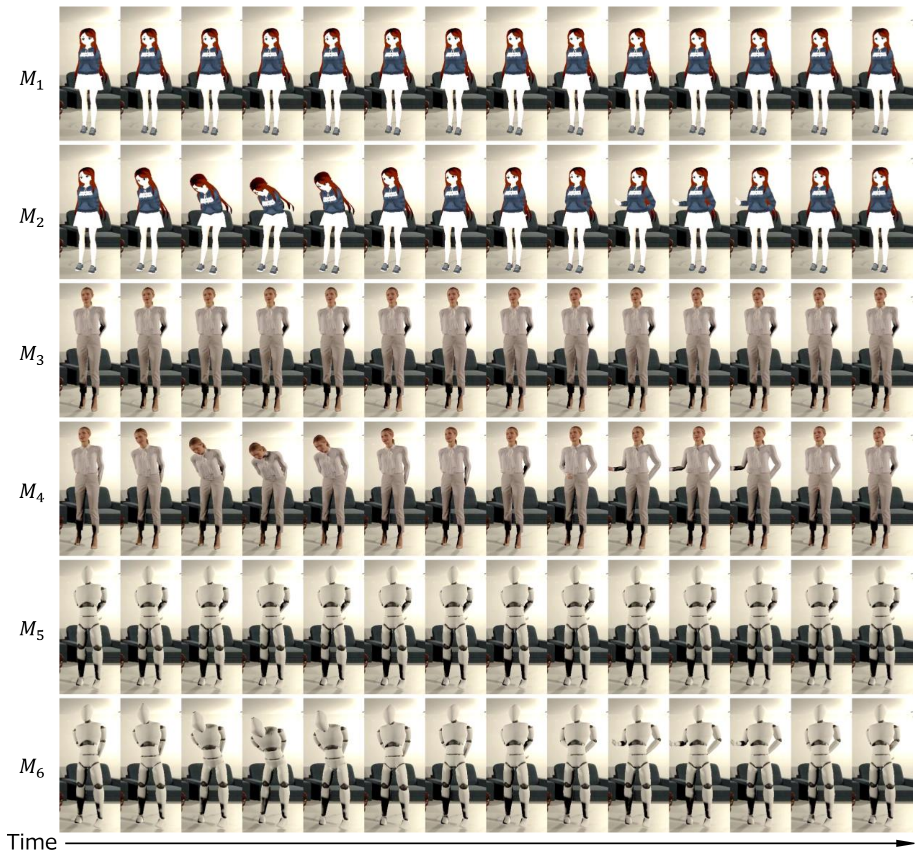

Fig. 3. Examples of the stimuli used during product description to test H_1 and H_2.

verbal content aligned with the intentions conveyed through the agent's gestures. For instance, the likelihood of the agent conveying "Welcome" and indicating a display product was exceedingly low. Incidentally, when the agent said "Please consider buying it," no gestures were performed.

2.5 Questions for the subjects

We recruited 22 participants (18 males, four females) for the experiment, with an average age of 21.5 years. We conducted subjective evaluations to validate H_1, H_2, and H_3. The questionnaire item was as follows:

Q_1**:** Did you perceive a good impression of the agent?

To ensure that subjects remained unaware of the experiment's intention, we introduced the following dummy item:

$\hat{Q}_1$**:** Did you experience any discomfort with the appearance of the product?

Fig. 4. Example of the red sphere before product description.

For each of the questions Q_1 and $\hat{Q}_1$, we formulated corresponding opposite items and presented a total of four questions to the subjects. The subjective evaluation values ranged from 1 ("Strongly Disagree") to 4 ("Strongly Agree"). The order of the questionnaire items presented to subjects was randomized across all subjects and comparison conditions. When presenting the questionnaire items, we did not disclose item names such as Q1 and $\hat{Q}_1$; only the text of the questions was provided. Opposite items were reverse-scored. The dummy items were not evaluated.

2.6 Procedure

The outline of the experimental procedures is as follows:

- P_1**:** Subjects put on the VR headset.
- P_2**:** We randomly choose one of the six comparison conditions.
- P_3**:** We present the red sphere to subjects and subjects observe the sphere.
- P_4**:** We present the selected stimulus to subjects.
- P_5**:** We instruct subjects to orally respond to the subjective evaluation questionnaire.
- P_6**:** We repeat steps P_2 to P_5 for each comparison condition.

The red sphere in P_3 is presented to ensure that at the beginning of the description, all subjects are looking at the same location for all comparison conditions. We displayed the red sphere to subjects at a position that intentionally does not overlap with the agent or the display product, as depicted in Fig. 4. To prevent surprising the subjects during the transition from P_4 to P_5, we issued the following verbal announcement: "We are now moving on to the questionnaire." During P_5, to ensure no impact on the results of the subjective evaluation questionnaire, we chose not to present the VR agent and display product to the subjects. Instead, we positioned the board with the questionnaire item written on it in front of the subject. They read the question and provided oral responses. If asked, the operator read the question aloud to the subject.

Before initiating the experiment, we explained the scenario to the subjects using three statements: "In this experiment, you work for a company," "You

S_1 S_2 S_3

Fig. 5. Three agent types with a display product.

have come to purchase a two-seater sofa planned for installation in you company's virtual space while inspecting the virtual product exhibition," and "You have the authority to make a purchase decision." These details were provided to immerse the subjects in a pseudo-situation in which they would receive information about a display product they are considering purchasing from the agent while simultaneously having the decision-making power for the purchase.

2.7 Other conditions

We chose a two-seater sofa as the display product for the agent to describe. Our inspiration comes from scenarios in which subjects explore physical stores for real-world display product purchases. In such situations, subjects often browse catalogs on web browsers for display products they might consider buying. However, potential discrepancies in color due to the environment and the risk of misinterpreting actual product sizes highlight the necessity for subjects to physically visit stores to verify aspects such as color and size. Therefore, in this study, we selected furniture items, specifically two-seater sofas, as the display product. Subjects might visit a store to directly assess details such as size. Figure 5 shows examples of the agent describing the sofa[1] in the stimulus. Figure 6 shows the spatial relationships among the subject, product, and agent.

When viewing the front of the display product, the subject is situated 0.5 m to the right of the product's center, whereas the agent is positioned 1.5 m to the right of the product's center. Similarly, when observing the left side of the display product, the subject is placed 1.5 m to the right of the product's center, and the agent is positioned 0.1 m to the right of the product's center.

2.8 Experimental setting

We used the VIVE Pro Eye (HTC Corp.) to display the VR space and mocopi (Sony Corp.) for motion capture. The mocopi motion capture system uses six sensors attached to the body, enabling 3D full-body tracking. In this study,

[1] We sourced the 3D data for sofas ZT8303DS and WT5603AS, as well as the light bulb LSJ-3_NK from Karimoku FreeBANK. The various tables and other sofas are from Digital-Architex.

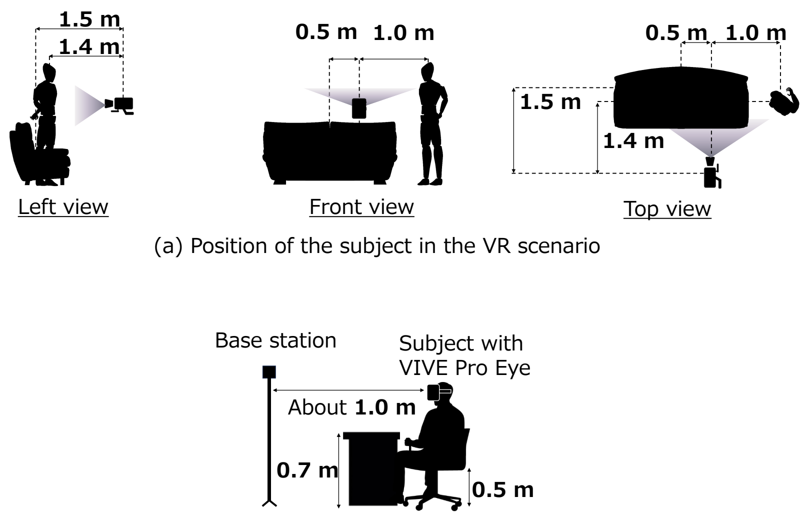

Fig. 6. Position of the subject with respect to the agent.

Table 1. Results of an analysis of variance on subjective scores indicating a positive impression towards the agent. * $p < .05$

Questuon	Variable factor	F-value	p-value	
Q1	Appearance(A_1, A_2, A_3)	4.03	0.02	*
	Gestures(G_1, G_2)	33.81	0.00	*
	Appearance(A_1, A_2, A_3) × gestures(G_1, G_2)	0.21	0.81	

our objective was to replicate gestures performed by a real-world clerk as they describe a display product to the subject. To achieve this, we captured the gestures of actual humans using motion capture and applied the acquired data to the VR agent with a humanoid skeleton. When obtaining the bowing gestures, we instructed the real-world human agent to imagine a subject was in front of them. Additionally, when obtaining the pointing gestures, we instructed the real-world human agent to imagine a display product diagonally to their right, as depicted in Figure 6. To create the audio data, we used the VOICEVOX: Haru-oto Ritsu voice synthesis software. We standardized the length of all stimuli to 10 s. The experimental setup, as shown in Figure 5(b), illustrates a subject sitting at the desk, adjusting themselves to a comfortable position for the experiment.

3 Subjective assessment of agent-described product presentation with gestures

3.1 Results of H_1 and H_2.

We conducted subjective evaluations to verify H_1, H_2 and H_3, obtaining subjective scores assessed by the experimental subjects. The Shapiro–Wilk test was

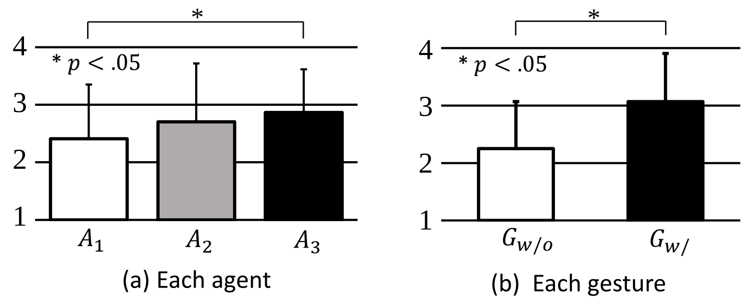

Fig. 7. Average subjective score for each agent or each gesture.

performed on the subjective scores rated by the subjects, and normality could not be assumed. Therefore, an aligned rank transform[5, 1] was applied, and an analysis of variance was performed on the appearance and gesture conditions of the VR agent. The result is listed in Table 1. A main effect was found for the VR agent's appearance ($F = 4.03, p < .02$). A main effect was also found for the VR agent's gestures ($F = 33.81, p < .00$). There was no interaction ($F = 0.211, p < .81$). Because there was a main effect for the VR agent's appearance, we believe that impression given by the VR agent changed depending on its appearance. Therefore, H_1 clearly holds. Because there was also a main effect for the VR agent's gestures, the impression given by the VR agent changed depending on its gestures. Therefore, H_2 also holds. Moreover, because there was no interaction between the VR agent's appearance and the VR agent's gestures, H_3 did not hold.

The average subjective score for each VR agent is shown in Fig. 7(a). The VR agent with the highest subjective score was A_3, followed by A_2, and finally A_1. Next, a Wilcoxon signed rank test was performed as a multiple comparison, followed by Bonferroni's correction. The results showed a significant difference between A_1 and A_3 (F=0.334, p<.045). These results indicate that the subjects had a better impression of the VR agent describing the product when it had a CG human-like appearance than when it had a robot-like appearance. The average subjective score for each gesture is shown in Fig. 7(b). The VR agent achieved higher average subjective scores of 3.07±0.84 for the presence of gestures and 2.25±0.82 for the absence of gestures. The order of higher average subjective scores was observed when actions were present compared with the findings when actions were absent. In conclusion, the agent achieved better outcomes when performing gestures while describing the display product.

3.2 Discussion

After analyzing the results of our hypothesis testing, we delve into the considerations arising from the subjective scores. This study focused on the influence of bowing and hand pointing gestures on the subjective scores of the agents.

Fig. 8. Examples of stimuli during a display product description with either bowing or hand pointing

The evaluation of H_1 revealed that an active engagement of the agent in gestures during product description resulted in a better impression from subjects. However, our investigation at this juncture has not clarified whether an effect was present when the agent executed either a bow or hand gesture individually, or whether the combined execution of both gestures had a discernible impact. As explained in Section 2.3, a bowing gesture expresses respect for the subject, and a hand pointing gesture conveys the direction of the product. When two different gestures are combined, the strength of the good impression that the user perceives may vary dynamically. We hence introduce a new hypothesis using the CG agents, which gave the best impression to the subjects, as shown in Section 2.3.

H_4 : The VR agent's bowing gesture changes the strength of the good impression that the user perceives.
H_5 : The VR agent's hand pointing gesture changes the strength of the good impression that the user perceives.
H_6 : Combining the VR agent's bowing gesture and hand pointing gesture changes the strength of the good impression that the user perceives.

3.3 Impact of each behavior on the subjective scores

We conducted experiments in which the agent exclusively engaged in either bowing or hand pointing gestures during product description, and we investigated how these gestures impact the subjects' perceptions of the agent by comparing them with the conditions outlined in Section 2.4, where gestures are either absent or present. The agent's bowing gesture conditions are as follows:

B_1**:** Without bowing
B_2**:** With bowing

The agent's pointing gesture conditions are as follows:

P_1**:** Without pointing

Table 2. Results of an analysis of variance on the subjective scores representing positive impressions towards the CG agent. * $p < .05$

Questuon	Variable factor	F-value	p-value
Q1	Bowing(B_1, B_2)	6.286	0.0140 *
	Hand pointing(P_1, P_2)	4.010	0.0460 *
	Bowing(B_1, B_2) × Hand pointing(P_1, P_2)	0.314	0.577

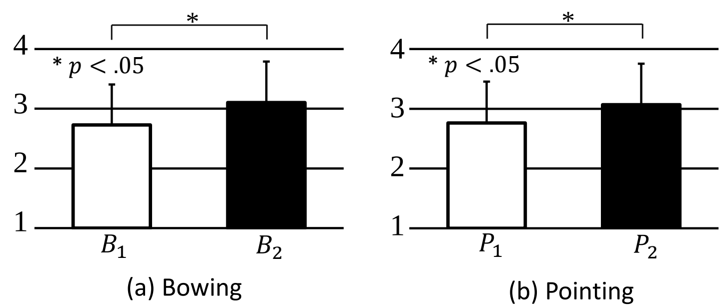

Fig. 9. Average subjective score for each bowing condition or pointing condition.

P_2**:** With pointing

We introduced the following four comparison conditions by combining the agent's bowing gesture conditions and the agent's pointing gesture conditions:

M'_1**:** CG human without bowing and pointing
M'_2**:** CG human with bowing
M'_3**:** CG human with pointing
M'_4**:** CG human with gestures

We provide the examples of stimuli for the M'_2 and M'_3 conditions in Fig. 8. Those for the M'_1 and M'_4 conditions are shown in Fig. 3 M_1 and M_2. The remaining experimental conditions and procedures follow those detailed in Section 2.4.

We conducted subjective evaluations to verify H_4, H_5, and H_6, obtaining subjective scores assessed by the experimental subjects. The analysis of variance described in Section 3.1 was performed on the bowing and hand pointing gesture conditions of the VR agent. The result is listed in Table 2. A main effect was found for the VR agent's bowing gesture ($F = 6.286, p < .0140$). A main effect was also found for the VR agent's hand pointing gesture ($F = 4.010, p < .0460$). There was no interaction ($F = 0.211, p < .0460$). Because there was a main effect for the VR agent's bowing, we believe that impression given by the VR agent changed depending on its bowing gesture. Therefore, H_4 clearly holds. Because there was also a main effect for the VR agent's hand pointing gesture, the impression given by the VR agent changed depending on its hand pointing gesture. Therefore, H_5 also holds. Moreover, because there was no interaction between the VR agent's bowing gesture and the VR agent's hand pointing gesture, H_6 did not hold.

The average subjective score for each bowing gesture condition is shown in Fig. 9(a). The subjective score of bowing gesture condition B_2 was higher than that of B_1. This order of higher average subjective scores was observed when bowing gestures were present as opposed to when they were absent. The average subjective score for each pointing gesture condition is shown in Fig. 9(b). The subjective score of hand pointing gesture condition P_2 was higher than that of P_1. This order of higher average subjective scores was observed when hand pointing gestures were present compared with when they were absent. In conclusion, when the agent described the product, even if the agent performed only one gesture, either bowing or hand pointing, the subjects perceived a better impression than they did when no gestures were performed.

4 Conclusion

We investigated three hypotheses to confirm whether experimental subjects' impressions of an agent describing a VR display product would be affected by the agent's appearance and gestures. The agent's appearance included robot, realistic human, and CG human conditions during the product description, with bowing and hand pointing gestures categorized into "without gestures" or "with gestures" conditions. We evaluated the six stimuli, combining the three appearances and two gesture conditions in a VR environment. Experimental results showed that subjects' impressions changed based on both appearance and gestures. Additionally, subjects favored the CG human agent over the robot agent. In an additional experiment, we investigated the variation in impressions when the agent performed a single gesture, either bowing or hand pointing. The findings indicated a positive impression even with just one of these gestures. In future work, we intend to conduct further analysis of the agent's appearance and an in-depth investigation of its gestures.

References

1. Elkin, L.A., Kay, M., Higgins, J.J., Wobbrock, J.O.: An aligned rank transform procedure for multifactor contrast tests. In: The 34th Annual ACM Symposium on User Interface Software and Technology. p. 754–768. Association for Computing Machinery (2021)
2. Hanus, M.D., Fox, J.: Persuasive avatars: The effects of customizing a virtual salesperson's appearance on brand liking and purchase intentions. International Journal of Human-Computer Studies **84**, 33–40 (2015)
3. Jang Ho Moon, Eunice Kim, S.M.C., Sung, Y.: Keep the social in social media: The role of social interaction in avatar-based virtual shopping. Journal of Interactive Advertising **13**(1), 14–26 (2013)
4. Liew, T.W., Tan, S.M.: Exploring the effects of specialist versus generalist embodied virtual agents in a multi-product category online store. Telematics and Informatics **35**(1), 122–135 (2018)

5. Wobbrock, J.O., Findlater, L., Gergle, D., Higgins, J.J.: The aligned rank transform for nonparametric factorial analyses using only anova procedures. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 143–146. Association for Computing Machinery (2011)