International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2017)
R. Lindeman, G. Bruder, and D. Iwai (Editors)

User Interface Agents for Guiding Interaction with Augmented Virtual Mirrors

Gun Lee[1,2], Omprakash Rudhru[2], Hye Sun Park[3], Ho Won Kim[3] and Mark Billinghurst[1,2]

[1]Empathic Computing Laboratory, School of ITMS, University of South Australia, Australia
[2]The Human Interface Technology Laboratory New Zealand, University of Canterbury, New Zealand
[3]Electronics and Telecommunications Research Institute, Republic of Korea

Abstract

This research investigates using user interface (UI) agents for guiding gesture based interaction with Augmented Virtual Mirrors. Compared to prior work in gesture interaction, where graphical symbols are used for guiding user interaction, we propose using UI agents. We explore two approaches for using UI agents: 1) using a UI agent as a delayed cursor and 2) using a UI agent as an interactive button. We conducted two user studies to evaluate the proposed designs. The results from the user studies show that UI agents are effective for guiding user interactions in a similar way as a traditional graphical user interface providing visual cues, while they are useful in emotionally engaging with users.

CCS Concepts

•***Human-centered computing*** → ***Mixed / augmented reality;*** *User interface design;*

1. Introduction

Augmented Virtual Mirror (AVM) systems provide an Augmented Reality (AR) [Azu97] experience on a large screen display with a video camera facing the user. This mimics a mirror like visualisation of the user standing in front of the system with virtual content overlaid on their image. Highlighting its potential usefulness in engaging with users, AVM displays have been applied to various applications such as showing virtual clothing [ES06], virtual shoes [MGT*07] [EFR08], or interactive entertainment [VGCF11]. Most of these prior work focused on the visual aspect of the interface, with only a few investigating interaction issues [LWP*15].

Gesture interaction is considered as one of the most natural ways of interacting with these type of displays, but it can be difficult to learn the gestures [QMB*02]. To overcome this limitation, there has been research on providing visual cues to help novice users learn gestures on desktop interfaces [BZW*09] [BM08], touch interaction [FBMW09] [DCN15], and motion recognition and training [SBW12] [WLF07] [AGMF13]. However there has been no prior work, to our best knowledge, on using user interface (UI) agents [Lau97] [LS03] as visual cues for guiding gesture interactions.

Compared to approaches explicitly instructing the users on poses and motions [PGI*17] [WBM13], we are interested in investigating how implicit cues [ABGB15] could be effective in guiding interactions, such as in physical objects with perceived affordance [Nor99]. This could be especially important in AVM applications as public information systems that need to attract novice users, and we propose that UI agents can be an effective way of engaging with users and guiding them through interaction. Our research focuses on how UI agents can guide novice users of an AVM system, and show them how to interact with the it.

2. Related Work

UI agents are intelligent autonomous virtual assistants that human users can collaborate with [Mae94] [Mae95] [Lie97]. Researchers have tried applying UI agents to various applications and tasks including web browsing [Lie95], presentation [RAM97], and education [LCK*97].

Natural speech and gesture interaction is considered the ultimate method for interacting with UI agents. For instance, Maes et al. [MDBP97] investigated using computer vision technique to recognize users' body motion to let them interact with a dog-like virtual agent, while Cassell et al. [CBC*01] supported more human-to-human communication like interaction with a humanoid interface agent.

However, to our best knowledge, there has been no prior work investigating how UI agents can be used to guide users how to interact with AVM displays using gesture. In the next section we describe a prototype system that uses a UI agent as a visual cue for guiding gesture interaction, and then a user evaluation conducted with the system.

DOI: 10.2312/egve.20171347

3. Interaction Design

We designed a prototype system to explore how UI agents could be used as guides for gesture interaction with AVM. We divided the user interaction with the AVM system into three stages: (1) Idle, (2) Engagement, and (3) Main application. In the Idle stage the system waits for a user to start using it. Once a user is identified, the system transitions into the Engagement stage where it helps the user to get ready for using the Main application. After the user finishes using the Main application and leaves, the system moves back into the Idle stage waiting for the next user.

The main role of the UI agent is to guide the user on how to interact with the system; guiding their attention to the place where to stand in the Engagement stage, and giving hints on how to select a button using gestures. We explain this in more detail in the following sections based on a simple quiz application to be answered by choosing a button with the correct answer.

3.1. Idle Stage

The idle stage is where the AVM is waiting for a user to come and engage with it. In this case face tracking is used to find and highlight the user's face as a cut-out on the advertisement image (see Figure 1), and the UI agent follows the potential user, trying to grab his/her attention. This shows the user that the system can track their motion and suggest using gestures to interact with it.

When a potential user is interested in the display they will turn towards it, move closer and look at the screen. To identify this, our system tracks the users shoulders to determine if they are facing towards the screen and if they are moving closer, and measure the time looking at the screen. We use an angle of about 15 - 20 degrees from directly facing towards the screen, and a trigger duration of three seconds. If the other requirements are met, during the three seconds the user is facing toward the screen the system enlarges the video hole, and transitions to the engagement stage.

Figure 1: *A UI agent catching potential user's attention.*

3.2. Engagement Stage

During the Engagement Stage the system guides the user where to stand to use the main application. A red outline of a pair of footprints is shown on the floor with a yellow animated arrow pointing to it (see left of Figure 2). To check if the user is standing at the right position, the AVM system measures the distances between the footprint symbol and the user's ankles. If the ankles are within 25 cm of the target footprints, and if they remain there for two seconds the system transitions to the Main Application Stage.

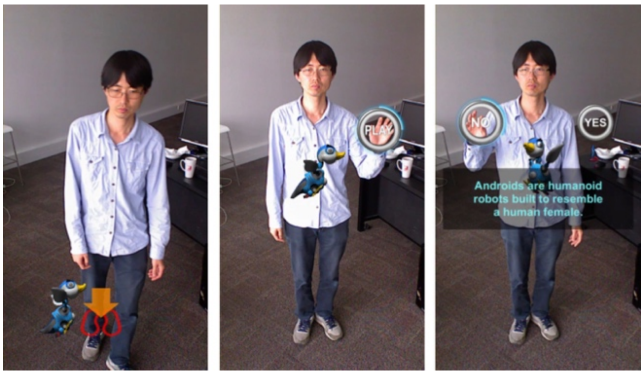

Figure 2: *A UI agent guiding user with interaction.*

3.3. Main Application Stage

The main application stage is where the user needs to actively interact with the system using hand gestures (e.g. selecting buttons). Here the UI agent guides users to select buttons on the screen using correct gestures. As a starting point, the system shows a 'Start' button which works as an introduction to how selecting buttons works (see middle of Figure 2). To indicate that the system is recognising the user's hand position, the UI agent follows around the user's hand. In case the system allows either hand, the UI agent follows the one that is raised higher.

To select a button, users have to hold their hand hovering on the button for a certain amount of time, which is a common gesture for button selection in an AVM [LWP*15]. As the user points at a button, the UI agent starts moving towards the button for selection. In order to suggest that the user has to hold his/her hand for certain amount of time, the UI agent adjusts its speed so that it will arrive at the button just after the user has held his/her hand on the button for enough amount of time. This design takes an analogy of the UI agent acting as a cursor with a certain amount of delay.

After few pilot trials, we learnt from user feedback that it is better to have the UI agent arrive a little bit earlier than when the button selection happens, and also to give certain feedback to help users to perceive as if the UI agent is selecting the button. In our prototype we had the UI agent arrive about a half a second earlier than when the button is selected, and wave its legs as if trying to push the button while making quack sounds.

4. Implementation

The prototype included a 55-inch TV screen mounted in portrait orientation on a mobile stand, with a Microsoft Kinect v1 sensor mounted at the top of the TV screen (see Figure 3). A computer (Intel Xeon E3-1240 3.4GHz CPU; 16GB main memory; and NVidia GeForce GTX 750 GPU) with Microsoft Windows 7 was used to run the prototype software developed in C# with the Unity 3D game engine.

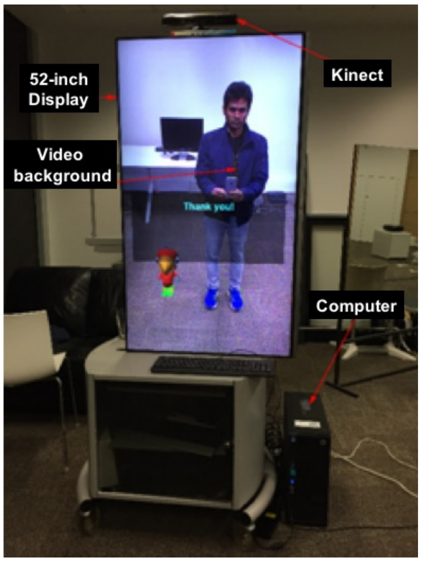

Figure 3: *Prototype system hardware setup.*

We implemented a custom plug-in software for Unity to stream colour (1280x720) and depth (640x480) images and user's skeletal tracking information using the Microsoft Kinect SDK. On screen, the prototype presented live video of the user's environment captured by the Kinect sensor in a virtual mirror configuration. Virtual content overlaid on the live video was registered using a direct linear transfer method [HZ03].

We implemented a behavior model of the UI agent to follow the user's hand as shown in Figure 4 where $\mathbf{A}_t$ represents the current agent position, and $\mathbf{T}$ represents the tracked hand position. Based on the two positions $\mathbf{A}_t$ and $\mathbf{T}$, we can calculate the agent position in the next frame $\mathbf{A}_{t+1}$:

$$\mathbf{A}_{t+1} = \mathbf{A}_t + s\mathbf{D}$$

where $s\mathbf{D}$ is a scaled vector of $\mathbf{D}$ (difference between $\mathbf{T}$ and $\mathbf{A}_t$) with the following scale factor s:

$$s = (1/fps)/t$$

where t is the preset amount of time delay in seconds for the UI agent to reach $\mathbf{T}$, and fps is the framerate (frames per second) of the system simulation/rendering loop. The time delay t is normally set to 1 second when the user hand is not on a button. When the user's hand is on a button, t is set to the amount of time left to hold the hand still until the button gets selected. In this way we can control the speed of the UI agent to let it arrive at the target position (on a button) just when the button gets selected.

We built a simple quiz application where users answered the questions by selecting 'Yes' or 'No' buttons. The theme of the quiz 'Birds and Robots' was chosen to match the interest and requirements of the project stakeholders. The theme was based on the historical background that the first known robot created around 400 - 350 BC by the mathematician Archytas was a steam powered pigeon. To match this theme the UI agent was presented as a robot bird.

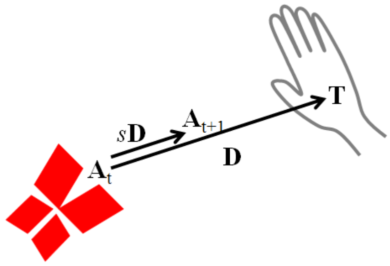

Figure 4: *UI agent behaviour of following a user's hand.*

5. User Study I

5.1. Study Design

To evaluate the prototype, we conducted a user study to investigate if the UI agent can effectively guide users how to interact with the AVM interface using gestures. Participants were asked to stand in front of the prototype system to do the experimental tasks and sit on a nearby chair for answering questionnaires, and debriefing. The main independent variable was the design of visual cues. We compared three types of visual cues: None (N), Graphical Symbol (G), and UI Agent (A) (Table 1). The experiment used a mixed design including both between-subject and within-subject designs applied in different parts of the procedure.

Table 1: *Visual cues provided in each condition.*

Condition	Visual Cues
None (N)	- Idle stage: Title image background - Engagement stage: Red footprint outline - Main Application: Button highlighted on hover and glows brighter on selection
Graphical Symbol (G)	In addition to the cues in condition N: - Idle stage: A hole of video stream highlighting user's face - Engagement stage: Animated yellow arrow pointing at the footprint - Main Application: Timer indicator around the button
UI Agent (A)	In addition to the cues in condition N: - Idle stage: UI agent following a hole of video stream highlighting user's face - Engagement stage: UI agent staying at the footprint - Main Application: UI agent following user's hand and selecting the buttons

We followed the procedure described in Table 2, which took about an hour for each participant without any restrictions on how long participants can interact with the system. The first session investigated how intuitive the system was hence not much explanation were given on how to interact with the display. Participants were asked to stand in front of the prototype and interact with it only based on what is shown on the screen. The participants experienced the Idle stage and the Engagement stage as the task was completed by selecting the Start button before the Main Application. Each participant experienced only one condition.

Table 2: *Experimental procedure with approximate duration.*

Duration	Procedure
5 min.	- Welcome - Informed consent - Pre-experiment questionnaire
Session 1 (between-subject design)	
10 min.	- Initial trial in one condition - Per-trial questionnaire - Interview
Session 2 (within-subject design)	
30 min.	Repeat for each condition: - Trial of the prototype system - Per-trial questionnaire
5 min.	Post-experiment questionnaire
5 min.	Interview & debriefing

The second session of the experimental trials was in a within-subject design, focusing on the button selection interaction in the Main application stage. The participants had three trials in different conditions. The order of conditions was counterbalanced using Balanced Latin Square design. In each trial, the task started from selecting the Start button followed by the Main application stage. Participants were instructed to rely on the information provided on the screen on how to interact with the system. Ten quiz questions on the topic of birds and robots were asked in each trial which the participants had to answer by selecting the Yes or No buttons. After each trial, participants were asked to answer the per-trial questionnaire, and after all three conditions, they answered the post-experiment questionnaire, and were interviewed briefly.

Per-trial questionnaires included the Brooke's System Usability Scale (SUS) [Bro96] and O'Brien's Engagement questionnaires [OT10]. The post-experiment questionnaire included ranking the conditions based on preference, and open questions to explain the reason behind the choice and giving feedback on how to improve the system. Through observation, we collected the number of times the wrong gestures were tried for selecting a button, and if participants smiled during the trial.

5.2. Participants

We recruited 21 participants, aged from 19 to 34 years old ($Mean = 26.2, Std.Dev. = 4.36, Median = 25$) and 10 of them were female (47.6%). Most of the participants were right hand dominant ($N =$ 18, 85.7%), the rest left. The observation and interview results are from 18 participants, as the records for the first three participants were not available due to technical issues.

The participants were not very familiar with gesture based interfaces. When asked if they had used gesture interface before, most of the participants ($N = 19$, 90.5%) answered less than few times a year or not at all. When asked if they had played Microsoft XBOX Kinect motion games before, 7 of the participants (33.3%) answered they had not played at all, while more than half of the participants ($N = 11$, 52.4%) answered a few times a year and only 3 of the participants ($N = 3$, 14.3%) answered a few times a week. When asked if they had used an AR app or interface before, 5 of them (23.8%) answered they were not aware of what AR is, while the rest ($N = 16$, 72.8%) had used at least few times a year (the mode was 'few times a month' as 7 participants has chosen this answer).

5.3. Results

The prototype was considered easy to use and figure out what to do from the first use. When asked if it was easy to know how to select a button rating in a scale of 0 (totally agree) to 10 (totally disagree), participants in the None (N) condition gave relatively lower rating $Median$ = 7.5 [IQR = 6-9], while the other conditions were rated 9 [8-10] and 9 [7-10] for the Graphical Symbol (G) and UI Agent (A) conditions, respectively. This is in line with the observations where two of the participants in condition N had troubles such as trying to touch the screen or trying to use the head to select the button. Also, more participants in the condition N were observed trying gestures other than hovering and waiting to select a button. About half (3 out of 6) of participants in both the G and A conditions tried tapping or pointing for selection, while all 6 out of 6 participants in condition N tried various gestures other than waiting such as, tapping, double tapping, and pushing. In the interview, when asked to explain how to select a button, only half of the participants in condition N described it in the correct way.

When asked about the relationship between the character and button selection in the final interview, about two thirds of participants (13 out of 18) described the relationship as the button gets selected when the UI agent arrives at the button. The two other participants did acknowledged the UI agent following the user's hand but were not sure about button getting selected, and the remaining four replied there is no relationship between them.

The SUS questionnaire results in session 2 showed all three conditions having slightly above the average (68) usability level (N: $Md = 80$, G: $Md = 82$; A: $Md = 77.5$), however no significant difference was found between the conditions ($\chi^2(2) = 3.552$, $p = .169$). The results of the Engagement questionnaire in session 1 trials showed all three conditions had slightly better than moderate level of engagement. The average scores of each condition were N: 3.52 ($Md = 3.68$), G: 3.31 ($Md = 3.2$), and A: 3.67 ($Md = 3.58$). Kruskal-Wallis test showed no significant difference between the conditions ($H = 2.770$, $p = .250$). The results of the session 2 were similar, where again all three conditions show slightly better than average level of engagement (N: $M = 3.73$, $Md = 3.84$; G: $M = 3.87$, $Md = 3.87$; A: $M = 3.7$, $Md = 3.84$). Friedman test found no significant difference ($\chi^2(2) = 2.375$, $p = .305$).

While no statistically significant difference was found in Engagement questionnaire results, observations and interview revealed a positive effect of the UI agent on user engagement. The researcher carefully observed the participants and made note of their behaviors. We found those who tried the UI Agent condition smiled more when using the system (see Figure 5). During the first session, while all the participants (6 out of 6) in condition A smiled, less than half of the participants smiled in the other conditions (N: 3 out of 6; G: 2 out of 6). More evidences of engagement with the UI agent was also found in the second section of trials where a few participants were found smiling when in condition A or even trying to play with the UI agent.

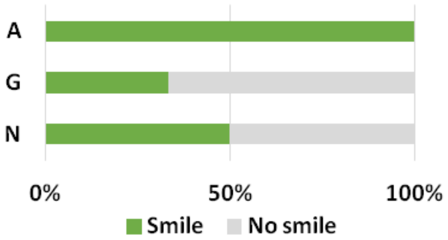

Figure 5: *Results of observation on participants smiling.*

Participants ranked the three conditions based on their preference (Figure 6). Most of the participants ($N = 18$, 86%) ranked the condition G in the first place, while more than half of the participants ($N = 12$, 57%) ranked condition N the last, and condition A was evenly split between the second and the third place while a few ranked it in the first place. A Friedman test found a significant difference in these rankings ($\chi^2(2) = 24$, $p < .0001$). Post hoc tests with Wilcoxon Signed Rank tests with Bonferroni correction ($\alpha = .016$) found a significant difference between condition G and the others (vs. N: $W = -231$, $p < .0001$; vs. A: $W = -192$, $p = .001$), but no significant difference between N and A ($W = -60$, $p = .259$).

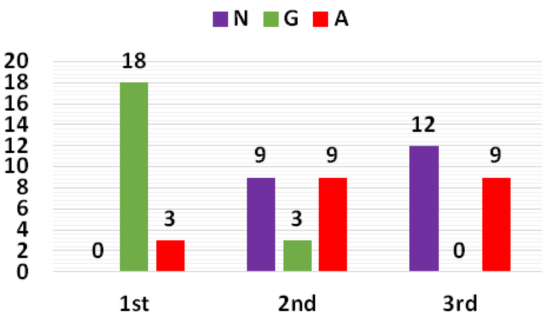

Figure 6: *Ranking results (number of participants).*

5.4. Discussion

About two thirds of participants described the relationship between the character and button selection correctly, showing that the UI agent was perceived as a good indicator of when the button gets selected. The number of participants having problems in condition A was on par with that in condition G, as they both reduced the errors to about half of that committed in condition N.

The results of the SUS and Engagement questionnaires showed no statistically significant difference between the conditions. This was probably because each condition has its own strength and weakness. The ranking results showed participants ranked condition G as the most preferred condition. In the interview they explained they felt it gave clear feedback on the time to wait when selecting buttons.

The main benefit of using the UI agent appeared to be from the emotional perspective of the user experience. This was demonstrated by all of the participants in condition A showing a smile on their face when trying the prototype for the first time. The character used in the study was not considered attractive based on interview results, but there were a few participants that still tried to play with the character. This leads to a conclusion that the UI agents could be beneficial to certain types of application that requires user's affection and engagement.

6. Alternative Designs of Interacting with UI Agents

Based on the feedback collected from participants in the first user study, we developed a second prototype system with improvements such as adding visual occlusion cues between the user's body and virtual objects and improving the design of the UI agent character to be more visually attractive (see Figure 7). In the engagement stage, we made the UI agent character to land on the ground and stand at a marked region of its own, suggesting the user should also stand at where it is marked for him or her. The marked region for the UI agent changed colour from red to green when it stood on it, suggesting the user is expected to stand on the larger region marked in red.

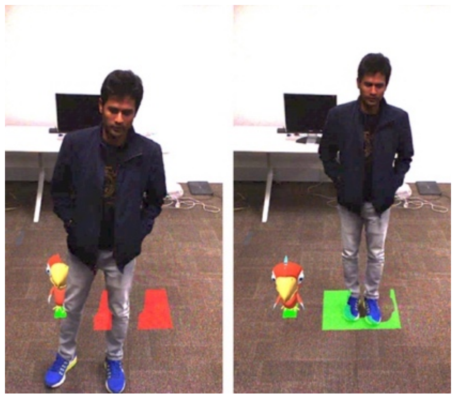

Figure 7: *Visual occlusion between the user and virtual objects.*

Apart from these minor improvements, the biggest change came from investigating alternative designs to make the relationship between the UI agent and the button more obvious. We initially explored the idea of changing the design of the button to look like an object that attracts the UI agent. For example, depending on the main application theme, if the UI agent is a bird character the buttons could look like its food, or if the UI agent has a form of a butterfly the buttons could be designed as flowers.

Further exploring with the idea, we developed an alternative design of interacting with the UI agent in which the UI agent plays the role of a button instead of a cursor, opposite to how it worked in the first prototype system. We hypothesised this could potentially solve the problem in which some of the participants in the first study felt that the UI agent is reacting slower with button selection compared to the graphical timer indicator around the button.

Figure 8 shows how this is applied to our second prototype using a quiz application asking about fruits and vegetables. The user is shown an image of a fruit or a vegetable, and he or she must answer if it is a fruit or vegetable by selecting the corresponding button. With the UI agent playing the role of the buttons, we used two bird characters with a thought bubble filled with fruits or vegetables, depicting what each bird is expecting. A smaller version of the fruit or vegetable image in question is used as a cursor following the user's hand, and as the user places his or her hand on one of the UI agents the bird charter showed an eating animation with the duration of how long the user must keep his or her hand on to confirm the selection. If the user has selected the correct answer the bird character showed an animation of flying, while for incorrect answers it showed an animation of falling down.

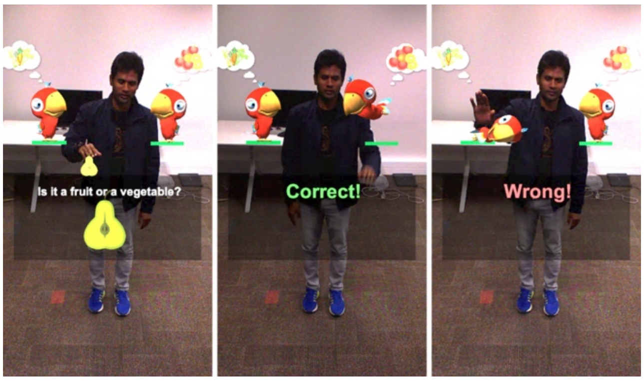

Figure 8: *UI Agents as buttons.*

7. User Study II

7.1. Study Design

A second user study was conducted to evaluate the revised prototype system. Two conditions were compared: Graphical Symbol (G) and UI Agent (A). The None condition was excluded as it was concluded being inferior to the other two conditions from the first study. The study procedure was mostly identical to the first user study.

7.2. Participants

We recruited 24 participants, none of the participants from previous experiment were chosen to participate in this experiment to avoid previous knowledge. The participants were between 18 to 63 years old ($M = 24.3$, $S.D. = 9.49$, $Md = 21$) and 9 of them were female (37.5%). Most of the participants used the right hand as their dominant hand ($N = 18$, 75%) while four participants answered that they use both hands, and two participants used the left hand.

The participants were not very familiar with gesture based interfaces. When asked if they had used gesture interface before, most of the participants ($N = 21$, 87.5%) answered less than a few times a year or not at all. When asked if they had played Microsoft XBOX Kinect motion games before, 9 of the participants (37.5%) answered they had not played at all, while more than half of the participants ($N = 12$, 50%) answered a few times a year. When asked if they had used an AR app or interface before, 8 of them (33.3%) answered they were not aware of what AR was, while the rest ($N = 16$, 66.6%) had used AR at least few times a year.

7.3. Results

The SUS questionnaire results showed both conditions having a slightly above the average (68) usability level (G: $Md = 81$; A: $Md = 75$), however a Wilcoxon Signed Rank test found no significant difference between the conditions ($Z = .4$, $p = .689$).

The Engagement questionnaire results showed that both conditions had slightly better than average level of engagement (G: $Md = 3.65$, A: $Md = 3.85$, see Figure 9). When compared with a middle value of 3, the Wilcoxon Signed Rank tests showed that both conditions were rated significantly higher than average (G: $Z = -4.286$, $p < .001$, A: $Z = -4.058$, $p < .001$). A Wilcoxon Signed Rank test showed there was a statistically significant difference between the two conditions ($Z = -2.334$, $p = .020$), with the UI agent condition having slightly higher rating.

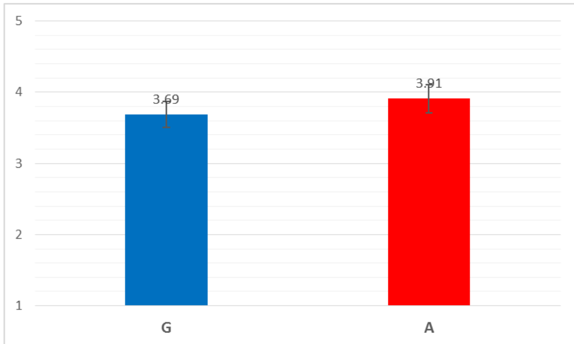

Figure 9: *Results of Engagement questionnaire.*

Figure 10 shows the results of the Engagement questionnaire broken down into each subscale. The average scores for the Focus subscale of user engagement for each condition were 3.13 ($Md = 3.0$) for condition G, and 3.4 ($Md = 3.5$) for condition A. A Wilcoxon Signed Ranks Test showed there was a significant difference between the conditions ($Z = -2.334$, $p = .020$). Similarly, significant differences were found for Novelty ($Z = -2.72$, $p = .006$), Endurability ($Z = -1.89$, $p = .058$) and Aesthetics ($Z = -2.15$, $p = .032$). No significant difference was found in the Involvement ($Z = -1.040$, $p = .020$) and Usability subscales ($Z = -.061$, $p = .951$).

In the post-experiment questionnaire, the participants were asked to rank the two conditions based on their preference. The results showed that more participants preferred the UI agent condition ($N = 14$, 58.3%; see Figure 11).

During the experiment, the researcher observed the participants to see if they were having any trouble with trying the wrong gestures to select a button. Wrong gestures appeared in the condition

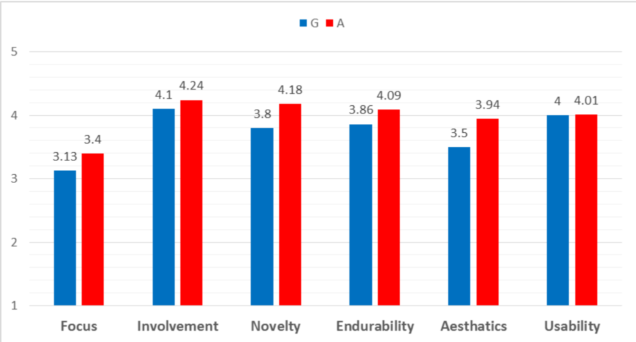

Figure 10: *Engagement questionnaire subscales.*

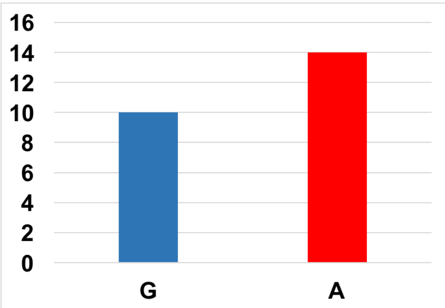

Figure 11: *Ranking results (number of participants).*

G included tapping the button (5), insufficient holding times (4), waving or swiping (3), pointing their fingers (2), and grabbing the button (1). On the other hand, errors committed in condition A included hovering their hand over the UI agent for an insufficient duration (3) and hovering their hand over the thought bubbles instead of the main body of UI agents (1). Participants made fewer errors in condition A ($M = 0.25$) than in condition G ($M = 0.67$). As a Shapiro-Wilk test indicated data of both conditions were not normally distributed (G: $W = 0.717$, $p < .0001$, A: $W = 0.531$, $p < .0001$), we conducted a Wilcoxon Signed Rank test found a significant difference ($Z = -2.178$, $p = .029$) between the two conditions, indicating participants in condition G made more than twice of the amount of errors compared to condition A.

Interview results indicated that while both conditions required the same amount of time (two seconds) to wait for selection after hovering over a button, many participants (17 out of 24, 70.8%) thought that the waiting time was different between conditions. More participants (10 out of 17, 58.8%) answered that condition A required less waiting time to select an answer, whereas the other participants believed that condition G required less waiting time to select the answer.

When asked to pick an interface that was fun, the majority of participants (18 out of 24, 75%) said that they had more fun while using the interface with the UI agents.

7.4. Discussion

Overall the second prototype system was accepted as very easy to use even in the first encounter. User engagement was higher in condition A with significant difference compared to condition G. There was a significantly higher user engagement for the subscales of Focus, Novelty, Endurability, and Aesthetics for condition A as compared to condition G. There was no significant difference in the usability subscale, which was in line with the SUS results.

Participants' preference for the UI agent interface was also reflected in terms of ranking for fun. More than 90% of the participants ranked the UI agent interface as the most fun interface and the majority of the participants rated it as their preferred interface. Some participants preferred condition G due to its simplicity, and some participants who preferred condition A liked it for its feedback and fun factor.

Interestingly, although both conditions required the same amount of selection confirmation time of two seconds, participants had varying perceptions of the amount of time they felt they had waited. More of the participants believed that the UI Agent condition required less waiting time than the Graphical Symbol condition. This could be due to the participants having more fun with a UI agent, which led them to believe less time had passed. Another reason for this could be that there were more visual elements to focus on which removed boredom whilst holding up their hands.

The users also made significantly fewer errors in condition A, trying incorrect gestures less. More incorrect gestures were made in condition G than in condition A.

Overall, the use of UI agents in an interface was found to improve user engagement in terms of focus, novelty, endurability, and aesthetics, as well as increasing the fun factor and preference of users. The UI agents also showed an improvement in the accuracy of the performed gestures, which could demonstrate the improved interpretation of the visual affordances. Due to the higher level of user engagement and fun factor, UI agents are considered to be more suitable in education and entertainment applications.

8. Conclusion and Future Work

We reported about our investigation on using a UI agent as interaction guides in AVM systems. The results from two user studies suggest UI agents are as effective at guiding users as visual cue in a traditional graphical user interface, while UI agents could be more beneficial for increasing the user's emotional engagement.

Future research includes improving the system design through applying the findings from this research, and also further investigation in how UI agents could guide gesture interactions other than button selection. We mostly focused on how UI agents can improve the user experience and engagement, yet further investigation would be needed on how they affect task performance. Also, further investigation through applying the UI agents to various applications would be necessary to generalise findings in this work.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.I5501-17-1016, Instant 3D object based Join&Joy content technology supporting simultaneous participation of users in remote places and enabling realistic experience).

References

[ABGB15] ALT F., BULLING A., GRAVANIS G., BUSCHEK D.: Gravityspot: guiding users in front of public displays using on-screen visual cues. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (2015), ACM, pp. 47–56. 1

[AGMF13] ANDERSON F., GROSSMAN T., MATEJKA J., FITZMAURICE G.: Youmove: enhancing movement training with an augmented reality mirror. In *Proceedings of the 26th annual ACM symposium on User interface software and technology* (2013), ACM, pp. 311–320. 1

[Azu97] AZUMA R. T.: A survey of augmented reality. *Presence: Teleoperators and virtual environments 6*, 4 (1997), 355–385. 1

[BM08] BAU O., MACKAY W. E.: Octopocus: a dynamic guide for learning gesture-based command sets. In *Proceedings of the 21st annual ACM symposium on User interface software and technology* (2008), ACM, pp. 37–46. 1

[Bro96] BROOKE J.: Sus: A âĂIJquick and dirtyâĂİ usability scale. usability evaluation in industry. edited by: Jordan pw, thomas b, weerdmeester ba, mcclelland il. 1996, 1996. 4

[BZW*09] BRAGDON A., ZELEZNIK R., WILLIAMSON B., MILLER T., LAVIOLA JR J. J.: Gesturebar: improving the approachability of gesture-based interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2009), ACM, pp. 2269–2278. 1

[CBC*01] CASSELL J., BICKMORE T., CAMPBELL L., VILHJALMSSON H., YAN H.: More than just a pretty face: conversational protocols and the affordances of embodiment. *Knowledge-based systems 14*, 1 (2001), 55–64. 1

[DCN15] DELAMARE W., COUTRIX C., NIGAY L.: A tool for optimizing the use of a large design space for gesture guiding systems. In *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (2015), ACM, pp. 238–241. 1

[EFR08] EISERT P., FECHTELER P., RURAINSKY J.: 3-d tracking of shoes for virtual mirror applications. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (2008), IEEE, pp. 1–6. 1

[ES06] EHARA J., SAITO H.: Texture overlay for virtual clothing based on pca of silhouettes. In *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality* (2006), IEEE Computer Society, pp. 139–142. 1

[FBMW09] FREEMAN D., BENKO H., MORRIS M. R., WIGDOR D.: Shadowguides: visualizations for in-situ learning of multi-touch and whole-hand gestures. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces* (2009), ACM, pp. 165–172. 1

[HZ03] HARTLEY R., ZISSERMAN A.: *Multiple view geometry in computer vision*. Cambridge university press, 2003. 3

[Lau97] LAUREL B.: Interface agents: Metaphors with character. *Human Values and the design of Computer Technology* (1997), 207–219. 1

[LCK*97] LESTER J. C., CONVERSE S. A., KAHLER S. E., BARLOW S. T., STONE B. A., BHOGAL R. S.: The persona effect: affective impact of animated pedagogical agents. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems* (1997), ACM, pp. 359–366. 1

[Lie95] LIEBERMAN H.: Letizia: An agent that assists web browsing. *IJCAI (1) 1995* (1995), 924–929. 1

[Lie97] LIEBERMAN H.: Autonomous interface agents. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems* (1997), ACM, pp. 67–74. 1

[LS03] LIEBERMAN H., SELKER T.: Agents for the user interface, 2003. 1

[LWP*15] LEE G. A., WONG J., PARK H. S., CHOI J. S., PARK C. J., BILLINGHURST M.: User defined gestures for augmented virtual mirrors: a guessability study. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (2015), ACM, pp. 959–964. 1, 2

[Mae94] MAES P.: Agents that reduce work and information overload. *Communications of the ACM 37*, 7 (1994), 30–40. 1

[Mae95] MAES P.: Artificial life meets entertainment: lifelike autonomous agents. *Communications of the ACM 38*, 11 (1995), 108–114. 1

[MDBP97] MAES P., DARRELL T., BLUMBERG B., PENTLAND A.: The alive system: Wireless, full-body interaction with autonomous agents. *Multimedia Systems 5*, 2 (1997), 105–112. 1

[MGT*07] MOTTURA S., GRECI L., TRAVAINI E., VIGANÒ G., SACCO M.: Magicmirror & footglove: a new system for the customized shoe try-on. *The future of product development* (2007), 441–450. 1

[Nor99] NORMAN D. A.: Affordance, conventions, and design. *interactions 6*, 3 (1999), 38–43. 1

[OT10] O'BRIEN H. L., TOMS E. G.: The development and evaluation of a survey to measure user engagement. *Journal of the Association for Information Science and Technology 61*, 1 (2010), 50–69. 4

[PGI*17] PUNPONGSANON P., GUY E., IWAI D., SATO K., BOUBEKEUR T.: Extended lazynav: Virtual 3d ground navigation for large displays and head-mounted displays. *IEEE transactions on visualization and computer graphics 23*, 8 (2017), 1952–1963. 1

[QMB*02] QUEK F., MCNEILL D., BRYLL R., DUNCAN S., MA X.-F., KIRBAS C., MCCULLOUGH K. E., ANSARI R.: Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction (TOCHI) 9*, 3 (2002), 171–193. 1

[RAM97] RIST T., ANDRÉ E., MÜLLER J.: Adding animated presentation agents to the interface. In *Proceedings of the 2nd international conference on Intelligent user interfaces* (1997), ACM, pp. 79–86. 1

[SBW12] SODHI R., BENKO H., WILSON A.: Lightguide: projected visualizations for hand movement guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012), ACM, pp. 179–188. 1

[VGCF11] VERA L., GIMENO J., COMA I., FERNÁNDEZ M.: Augmented mirror: interactive augmented reality system based on kinect. In *IFIP Conference on Human-Computer Interaction* (2011), Springer, pp. 483–486. 1

[WBM13] WALTER R., BAILLY G., MÜLLER J.: Strikeapose: revealing mid-air gestures on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), ACM, pp. 841–850. 1

[WLF07] WHITE S., LISTER L., FEINER S.: Visual hints for tangible gestures in augmented reality. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on* (2007), IEEE, pp. 47–50. 1