International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2019)
Y. Kakehi and A. Hiyama (Editors)

Evaluation of Embodied Agent Positioning and Moving Interfaces for an AR Virtual Guide

Nattaon Techasarntikul[1], Photchara Ratsamee[2], Jason Orlosky[2],
Tomohiro Mashita[2], Yuki Uranishi[2], Kiyoshi Kiyokawa[3], and Haruo Takemura[2]

[1]Graduate School of Information Science and Technology, Osaka University [2]Cybermedia Center, Osaka University
[3]Graduate School of Science and Technology, Nara Institute of Science and Technology

Abstract

Augmented Reality (AR) has become a popular technology in museums, and many venues now provide AR applications inside gallery spaces. To improve museum tour experiences, we have developed an embodied agent AR guide system that aims to explain multi-section detailed information hidden in the painting. In this paper, we investigate the effect of different types of guiding interfaces that use this type of embodied agent when explaining large scale artwork. Our interfaces include two types of guiding positions: inside and outside the artwork area, and two types of agent movements: teleporting and flying. To test these interfaces, we conducted a within-subjects experiment to test Inside-Teleport, Inside-Flying, Outside-Teleport, and Outside-Flying with 28 participants. Results indicated that although the Inside-Flying interface often obstructed the painting, most of the participants preferred this type since it was perceived as natural and helped users find corresponding art details more easily.

CCS Concepts

• ***Human-centered computing*** → *HCI design and evaluation methods;* • ***Computing methodologies*** → *Mixed / augmented reality ;*

1. Introduction

Museums and historical sites are places for leisure that also deliver both knowledge and entertainment to wide range of people. Typically, many museums have a tour guide for showing a group of visitors around while providing supplemental information about exhibits. Human guides have the ability to attract visitors by matching their presentation style to the current actions or intentions of visitors. By joining a tour with a guide, a visitor can both gain knowledge and be entertained. However, with the limited number of guides coupled with labor costs, museums cannot always provide a one-on-one customized experience to every visitor who comes to the museum. To help address this problem, audio and AR guided applications have become an alternative means to enhance the museum experience and also provide in-depth information of the exhibit to the visitor. While the audio guide provides only a description, AR guide applications have a more interactive component. These kinds of AR applications such as Blippar [AM11] and Layar [MLF09] are usually easy to create and maintain. Though Blippar and Layar typically only run on smartphones or tablet, they do provide one-to-one interactions in which users can explore an exhibition at their own pace.

In this paper, we introduce a wearable AR guide system that makes use of an embodied agent as a virtual guide who points out and explains each part of a painting to a visitor. Although much research about the embodied agents exists [CSHO14, AKYT00, BS04], it has not been focused on using an embodied agent as a companion guide for real world use. In this sense, our virtual guide acts in the same manner as a human guide, for example by standing next to an exhibit while describing its contents and maintaining eye contact with the visitor. Moreover, in addition to implementing both verbal and non-verbal communication interfaces, we also include a virtual representation of the embodied agent. This virtual existence enables many functions that are not possible for a human guide, such as showing another virtual image, flying around to reach a high position, or using teleporting functionality to reduce travel time to other real world spaces. Since such virtual functions have not yet been studied in the context of real world situations, we conducted a user study to evaluate the following Research Questions (RQ):

- RQ 1: Do users mind if the embodied agent position inside the painting overlaps some of the art?
- RQ 2: Does teleporting help reduce total guiding time?
- RQ 3: Are there any differences in memory retention between each method of guidance?

The goal of this study is to explore which virtual guide movement and positioning interface users prefer and to explore whether different interfaces affect the user's tour experience and memory retention. We expect that our AR agent guidance system will help the user concentrate and remember information while shortening the time used for exploring information in the painting.

DOI: 10.2312/egve.20191276

2. Related work

2.1. Augmented reality system for a museum

The study of AR mobile applications in museums started more than a decade ago. For example, Miyashita et al. [MMT*08] developed an AR Magiclens system and found some issues such as small text size that depends on the camera's distance to the artwork and an AR with a black color background on the screen because of a dark environment. Some visitors also commented that it was difficult to switch their attention between an AR monitor and the real artwork while holding the monitor with both hands.

While a hand-held based AR system requires the user to hold it up to see the AR image, smart-glasses and AR displays are good alternative hands-free devices. Mason [Mas16] conducted a study using a Google Glass as an AR guide application. Many participants said it was easier to see a video description and the real object at the same time. However, some participants commented that a text description was too long to read, and they are not unfamiliar to read text shown in smart-glasses. He also found that the users only watch the video for around 10 to 20 seconds out of the full 70 second length and then switched their attention to the real object in front of them. This suggests that a user may not focus until the end of the description if it is not interesting enough.

Vainstein et al. [VKL16] have studied the user requirements of smart-glasses for museum visiting. Based on the results of a survey taken from a museum, most participants are preferred to control the AR smart-glass system using an accompanying mobile device, not by voice, due to the quiet environment of the museum. They also preferred a function to adjust text size, audio volume, fast forward or pausing of media, and automatic stopping of the presentation when the user moves away. Moreover, many of them did not prefer the system to display text because it was not easy to read.

Since we focus on developing a system that gives an explanation of each section of a large piece of artwork, neither a hand-held AR device nor AR smart-glass with a small field of view (FOV) would be able to overlay AR guidance covering the entire area of the work. The user needs to move his or her AR display to see each piece of guidance for the piece. In this situation, someone wearing smart-glasses could just turn his or her head to see AR guided information, which would be easier than raising a hand-held AR device to a new position. To minimize the need for hand movement but maintain the advantage of adequate tracking, we decided to develop our system on a wearable display and have voice explanations with the embodied agent guide interface.

2.2. Embodied agent

An embodied agent is usually considered a sub-field of computer simulation, which aims to develop software that has a humanoid or body-like appearance and verbal and nonverbal communication similar to that of a human. There are embodied agents that are visualized in a monitor for weather reporting, that can walk around and refer to objects in a virtual environment [NZB00], language trainers [Mas04, MGR14], hospital companions [BAE*15], support for elderly individuals [RSTB15], and museum guides [KGKW05, STA*10]. These embodied agents were rated as effective and can often be used in place of a humans who are not available for 24 hours a day. Some embodied agents are display in AR setting environment that a user will see the agent integrated to the real world seamlessly by seeing through a head-mounted display, which it helps a user in furniture arrangement [AKYT00], guide a user in factory and demonstrate how to use a machine [VLP*03].

Regarding comparisons of the effectiveness of embodied agents, Bickmore et al. [BPY08] have studied on user preference between a human and an embodied agent on a complex document explanation task. They found that users are satisfied on an explanation from the embodied agent than a professional human with the reason that the agent has infinite patience and repeat a word, and will not criticize if they don't understand something. Also, Beun et al. [BDVW03] have studied on effects on memory recall of the presence of an embodied agent. They found that participants who listened to the embodied agent's voice had a significantly higher memory recall score compared to others who listened to audio only. Moreover, Campbell et al. [CSHO14] conducted an AR navigation task and found that by following an embodied who leads a walking participant, navigation was faster than following arrows placed along the way.

When considering AR guidance development, adding an embodied agent to the system could bring many merits. Our embodied agent will act as a virtual guide who talks, smiles, pays attention to the user, and points out interesting parts inside the artwork to the user. We expected that our embodied agent would be preferred by the user, that the user would pay more attention to the information being explained, and that the whole experience would be improved with the AR guide in place.

3. Design and Technical Implementation

Our AR guide system is designed to explain all available regions of a large painting containing multiple details by pointing out each position of guidance and explaining the information pertaining to that piece (region of the image) to the user. We decided to develop this system using the Microsoft Hololens, an optical see-through head mounted display [Cor16], which is a standalone device that can perform real-time localization and visualize high-quality mixed reality images. Our system uses images from the device's forward-facing camera to track the position of the painting itself, and also receives button input from a Microsoft Xbox controller to trigger the embodied agent to explain subsequent points of information.

Due to the limited field of view of the device, the size of an embodied agent is resized to fit the display height, which prevents cutoff of the virtual guide image viewed by the user. Also, to able to see the agent face expression more easily, we decided to use an embodied agent type that has a slightly larger head proportion, the young female character model called SD-unitychan [Jap14], as our virtual guide. We set a height of the model to 50 cm, which fits the device display height when the user stands inside a proxemics based social distance [HBB*68], around 160 cm away from the embodied agent. This model is already rigged and has some basic facial animations provided so that we can program movement of the hands, mouth, and eyes. Our embodied agent contains three primary functions, including avatar communication, positioning, and movement.

Figure 1: Images of the experiment setup showing (a) the area in which the experiment took place and then the actual view through HoloLens with (b) Outside type and (c) Inside type position interfaces.

3.1. Avatar Communication

Both verbal and non-verbal means of communication are included in our embodied agent. A guide narrative is generated from commercial text-to-speech software [AI09]. For non-verbal communication, eye contact is an important means for creating a communicative relationship [IOI*01]. Therefore, our agent always makes eye contact with the user by facing its body towards the position of the user while describing the painting. This can make the user feel that the agent is talking to him/her. Also, our embodied agent uses an open hand gesture to point toward the explanation part and indicate the direction of the explanation. A single left or right hand is determined from the side closest to the position of the text. Based on our pilot study on pointing interfaces [TRO*19], we included a virtual line laser to explicitly point out the position that helps the user more quickly find the part of the painting about which the virtual guide is talking.

3.2. Avatar Positioning

When the user is wearing a HoloLens, a virtual graphic that is seen by the user is only in the area of 30-degrees horizontal and 17-degrees vertical in front of the head position, which does not cover all of the area seen by the user's eyes. If we put the SD-unitychan embodied agent on the floor, the users might not see the agent unless they looked downwards. Therefore, we decided to allow our embodied agent to float in the air so that the user can see the agent in their normal view. Since the embodied agent is floating freely in the air, it can overlap a painting while describing. To study if a painting overlapping has a negative effect on a guiding or not, we created 2 guiding position interfaces: position inside and position outside. Using the position outside interface, the embodied agent does not occlude the area of the painting by stays at one border of the painting at the same height of the explained segment. we calculated the agent position at both left and right border and select the side that closest to the explaining position. In contrast, the embodied agent with the position inside interface occludes some area of the painting but stays next to the explained segment. We assume that the users who pay attention to the virtual guide would prefer to have the embodied agent stay inside their view while listening to the talk.

3.3. Avatar Movement

When the embodied agent starts talking about the new subsection of the painting, a new position will be applied. To move the virtual guide from the current position to the new position, we provided 2 movement interfaces, teleporting and flying. The teleport interface moves the embodied agent to a new position immediately so that the user doesn't have to to wait for the agent to reach the new position and start explaining a new item. In contrast, the flying interface gradually moves the agent as a straight-line path to the new position, which takes time if the agent needs to move from one side of the artwork to another. It takes 2-3 seconds for the embodied agent to fly from one side of the painting to the other. By combining avatar positioning and movement, we have 4 guiding interfaces, which include: Outside-Flying (OF), Outside-Teleport (OT), Inside-Flying (IF), and Inside-Teleport (IT) to be used by our embodied agent guide.

The embodied agent guide with the OF interface does not occlude the painting while describing the corresponding piece of information in the painting. However, when the agent flies to the other side of the painting, it occludes the painting and the user must wait for the movement to be finished before listening to the next narration. The embodied agent guide with the OT interface does not occlude the painting while guiding a participant. When it needs to move to another location in the painting, it will disappear from its current position and appear at the new position. The user can turn his or her head and quickly find the new positions of the agent and explanation. The embodied agent guide with the IF or IT interface floats nearby the description while occluding other parts of the painting. With the IF interface, the agent will fly from its current position to the next position, which takes longer when compared to the IT type.

4. Evaluation

4.1. Experiment setup

The painting we used for testing is "The Netherlandish Proverbs," by Pieter Bruegel the Elder in 1559 [Wik09]. The painting is in the public domain, meaning it is free to use. We printed this image at 200 by 140 cm and hung it on a wall 100 cm above the floor. To

(a) First trial (b) Second trial (c) Third trial (d) Forth trial

Figure 2: The painting and detailed positions for explanation that were used in our evaluation of the system.

facilitate viewing, we provided a 250 by 260 cm free walk area in front of the painting as shown in Figure 1.

We then prepared 16 different subsections of the image, with sets of four sections grouped together to test the four guiding interfaces (OF, OT, IF, IT), the orders of which were balanced to alleviate learning effects. The guiding started in the same order, from the first trial to the forth trial, and in the same order in each trial from the most left circle to the most right circle as depicted by the directional arrow in Figure 2. An explanation for each part saying a location of that part in the painting, the image characteristic to look at, and the meaning of that image. The explanation audio ranged from 11 to 17 seconds, with an average of 14.38 seconds. One example of an explanation is "At the right side of the painting, you will see a man who is catching an eel on its slippery tail. This is called to hold an eel by the tail, which means to undertake a difficult task." Before the explanation start, the embodied agent will move to the position near the describing part with the selected positioning and movement interface. After the avatar reaches the determined position, it will point to the direction of an image and start talking.

In our experiment, we recruited 28 participants (13 female, 15 male, mean age of 21, SD of 1.61) to help evaluate the effectiveness of the guiding interfaces. Participants were mostly students, and came from various departments, studying in the Schools of Language, Letter, Human Sciences, Health Sciences, Biosciences, Law, Economics, Sciences, Engineering, and Engineering Sciences. The ratio of sciences and non-sciences students was 11:17. We paid them with a bookstore card valued at 500 yen (around 5 USD) for participating in the 20 minute experiment.

4.2. Procedure

First, we explained the experiment procedure, including a description of how to use the AR guide system to each participant. we also informed that there would be a test after the experiment and asked him/her sign a consent form. We then started up the application and let the participant stand 230 cm away from the center of the painting. At this position, the participant could see the overlaid graphics in a corresponding physical area of 130 * 70 cm.

Then we let him/her wear the HoloLens and confirm that he/she sees the embodied agent at the left corner of the painting which is a standby position. Next, we informed the participant that the experiment will start when he/she presses a button and he/she is free to move around the designated viewing area. The times when the user triggered explanations and standing positions in the environment were recorded during the experiment. The experiment started with demographic questions. Then, the virtual agent guided the 4 segments of the image. The participant had to press a button on a wireless controller to trigger the virtual guide to proceed to the next location. After participants finished listening to all four explanations present in the painting, they had to complete a subjective score related to each of the guiding interfaces. Items were scoured on a Likert scale for the following statements:

1. The virtual guide obstructed the area of explanation.
2. the laser pointer obstructed the area of explanation.
3. A position of the virtual guide was proper.
4. It was easy to find the area of explanation.
5. This interface looked natural.
6. This interface help me concentrate on the description.
7. This interface help me concentrate on the area of the explanation.

The guidance and subjective questionnaire were repeated four times for the four interfaces, selected randomly. Finally, we let the participant select one of his/her preferred guiding interfaces (from the four, OF OT IF and IT). Note that both the demographic questions and subjective questionnaire were embedded inside the HoloLens application so that the participant did not need to take off the device and put on again in the middle of the experiment.

When the participant finished the experiment and questionnaire, we then asked him/her to complete a memory test on a computer. The test consisted of an image of the painting and 16 text-boxes with a different phase place around the image. Each phase was referred to during the explanation dialogue corresponding to the meaning of part in each area. The participant was then asked to draw an arrow from each phrase to its corresponding position. For an example, he/she has to draw an arrow from a phase "To undertake a difficult task" to the position that related to the phase which is an image of a man who is catching an eel on the right side of the painting. Participants are allowed to take time as much they need or give up anytime. After that, we interview each one about the guide impression, why he/she prefer the selected positioning and movement interface, and how much he/she can do the test.

4.3. Experiment Results

We evaluated the effects of each interface on both objective and subjective measures. The objective measures included the time that

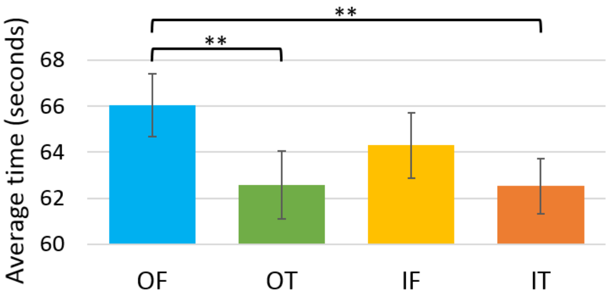

Figure 3: Average time that the participant used for experience each guiding interface including Outside-Flying (OF), Outside-Teleport (OT), Inside-Flying (IF), and Inside-Teleport (IT).

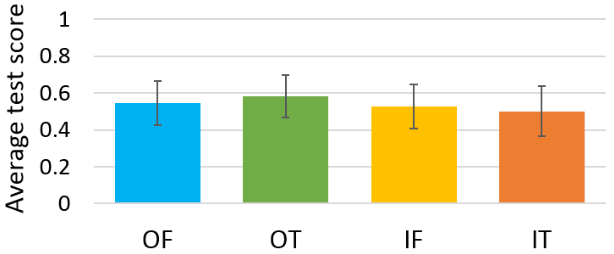

Figure 4: Average number of correct participant answers when they experienced each guiding interface.

participants spent on experiencing the AR guide system with each interface and a score from a memory recall test conducted after the participant used all 4 interfaces. We use a repeated measure Analysis of Variance (ANOVA) test [Gir92], following with a post-hoc analysis using pairwise t-tests [Efr69] with Holm's adjustment [Hol79] to analyze the objective measures.

We collected subjective measures by asking the participant to rate each interface regarding obstructiveness, appropriate position, ease of finding the corresponding explanation, naturalness, and assistance to concentration using 5-point Likert scale [AS07]. We used a non-parametric repeated measures Friedman test [SFT96], following with a post-hoc analysis using Wilcoxon signed-rank test [Woo07] with Holm's adjustment method. We indicate a significance level of $p < 0.001$, $p < 0.01$, and $p < 0.05$ using asterisks ***, **, and *, respectively, between each set of bars in our results plots and figures.

4.3.1. Time Used

The time for each guiding interface that each participant used to navigate the 4 parts (from the left to right side of the painting) is shown in Figure 3. Outside-Flying (OF) took the longest time, with an average of 66.04 seconds, followed by Inside-Flying (IF) at 64.29 seconds, Outside-Teleport (OT) at 62.58 seconds, and Inside-Teleport (IT) at 62.53 seconds. We first conducted an ANOVA and found a significant difference between interfaces, ($F = 5.99$, $p < 0.001$). We then conducted a post-hoc analysis and found that

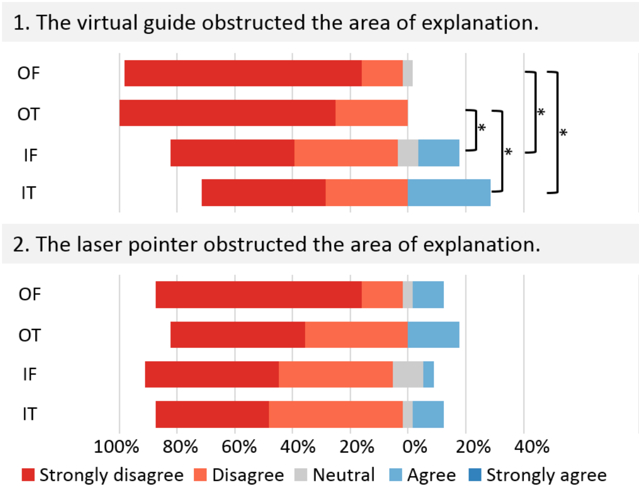

Figure 5: Subjective ratings for each interface regarding obstructiveness to the real painting.

mean time spent was significantly shorter with OT than with OF ($p < 0.0025$), and IT also required less time than OF ($p < 0.0025$).

4.3.2. Memory Recall

The number of correct matches between painting locations and descriptions varied from person to person. The average score for each group of descriptions and the difference of each interface are shown in Figure 4. When guiding with OT interface, participants answered the test corrected 58%, 54% with OF interface, 53% with IF interface, and 50% with IT interface. We conducted the ANOVA test but no significant difference was found between interfaces, ($F = 0.297$, $p = 0.827$).

4.3.3. Obstructiveness

In addition to objective measurements, the results of obstructiveness to the painting for the virtual guide and laser pointer for each interface are shown in Figure 5. With respect to obstruction, results showed that 28.57% of participants rated agree and 71.43% rated disagree on IT interface, whereas 14.29% rated agree and 78.57% rated disagree for the IF interface. None of them rated agree that the virtual guide with OF and OT interfaces were obstructive.

Regarding whether the laser pointer obstructed the area of explanation, 17.86% of participants rated agree and 82.14% rated disagree on OT interface. Both OF and IT interfaces 10.71% of participants rated agree and 85.71% rated disagree. For the IF interface, 3.57% of them rated agree and 85.71% rated disagree, while 10.71% rated neutral.

We ran a Friedman test and found a significant difference between methods for the virtual guide obstructiveness, $\chi^2(3) = 27.046$, $p < 0.001$. The post-hoc analysis result reveals the same level of significance ($p < 0.5$) between IF and OF, IF and OT, IT and OT, and IT and OF. However, no significant differences were found for the laser pointer obstructiveness, $\chi^2(3) = 6.252$, $p = 0.09$.

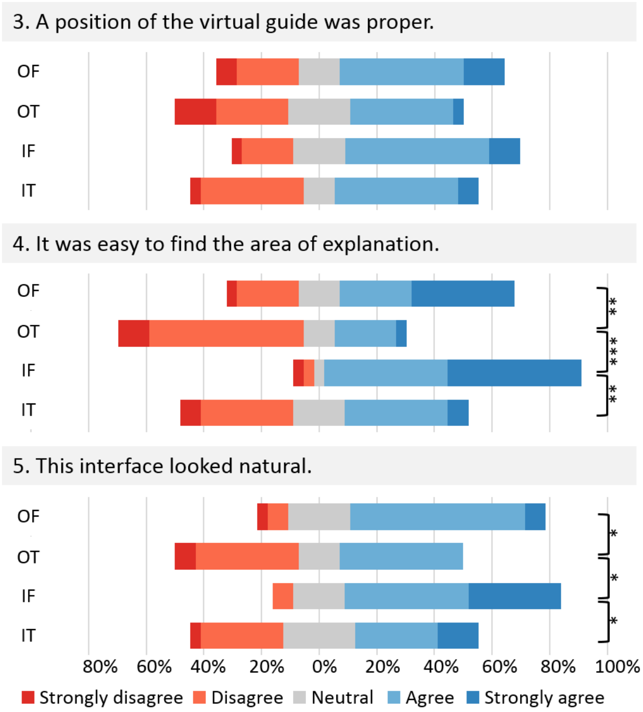

Figure 6: Subjective ratings for each interface regarding interface position, discovered position and naturalness including Outside-Flying (OF), Outside-Teleport (OT), Inside-Flying (IF), and Inside-Teleport (IT).

4.3.4. Interface Position, Ease of Finding, and Naturalness

Subjective results regarding appropriate position, ease of finding the explanation position, and naturalness of each interface are shown in Figure 6. Regarding a position of the virtual guide was proper or not, 60.71% of participants were agree and 21.43% were disagree on IF interface. With OF interface 57.14% of them were agree and 28.57% were disagree. With IT interface 50% of them were agree and 39.28% were disagree. While the same 39.28% of participants shared agree and not agree with OT interface.

Regarding the area of explanation was easy to find or not, 89.28% of participants were agree, while 7.14% of them were not with the IF interface. For the OF interface, 60.71% rated agree and 25% rated disagree. For the IT interface, the number of the participants rated agree and disagree are similar at 42.85% agreed and 39.28% disagreed. While 25% rated agree and 64.28% rated disagree to the OT interface.

Regarding the interface looked natural or not, 75% of participants rated agree and 7.14% rated disagree to the IF interface. With OF interface 67.86% of them rated agree and 10.71% rated disagree. While with the OT interface, both agree and disagree were rated equally 42.86% from the participants. For the IT interface, 42.86% agreed and 32.14% disagreed.

We ran a Friedman test and found significant differences on ease of finding and naturalness of each guiding interface with $\chi^2(3) = 28.913$, $p < 0.001$, and $\chi^2(3) = 14.325$, $p < 0.01$, respectively. How ever no significant difference found on the virtual guide position ($\chi^2(3) = 3.824$, $p = 0.281$).

The post-hoc analysis revealed significant differences in the rating of ease of finding, and OT had a lower score compared to OF ($p < 0.01$) and IF ($p < 0.001$). Also, the IF interface has a higher score compared to IT ($p < 0.01$). Together with the naturalness rating result, we found the same level $p < 0.05$ of significant differences between OT and OF, OT and IF, and between IF and IT interface.

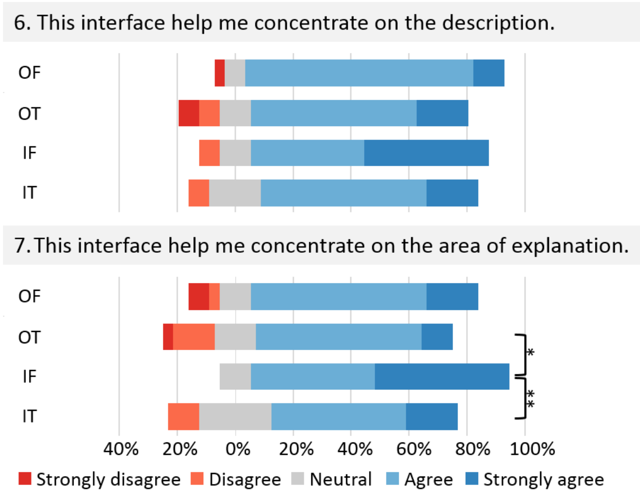

Figure 7: Subjective ratings for each interface regarding concentrate on the explanations and area in concern.

4.3.5. Concentration

Results of concentration on the content description and area of explanation effected by each interface are shown in Figure 7. Regarding concentration on the content description, 89.28% of participants were agree and 3.57% were disagree with the OF interface. For the IF interface 82.14% of them were agree and 7.14% of them were disagree. There are 75% of participants who agreed with OT and IT interfaces. While 14.28% and 7.14% disagree on OT and IT, respectively.

Regarding concentration on the area of explanation, 89.28% of participants agreed with IF interface none of them disagreed. For the OF interface, 78.57% agreed and 10.71% disagreed. For the OT interface, 67.85% agreed and 17.86% disagreed. For the IT interface, 64.28% agreed and 10.71% disagreed.

We ran a Friedman test and found a significant difference on area of explanation concentration ($\chi^2(3) = 11.751$, $p < 0.01$), but no significant effect on the content description concentration ($\chi^2(3) = 5.2241$, $p = 0.156$). The post-hoc analysis revealed significant effects on aiding concentration of the explanation area between IF and IT ($p < 0.01$) and between IF and OT ($p < 0.05$).

4.3.6. Interface Ratings

We asked the participants to choose the interface that they prefer most at the end of all questionnaires. Twelve participants (42.86%)

preferred IF interface, eleven (39.29%) preferred OF interface, three (10.71%) preferred IT and only two (7.14%) preferred the OT interface. Results shown in Figure 8 correspond to the previous subjective rating results that the most voted, IF and OF interface had many positive significant difference in several areas.

4.4. Discussion

In the evaluation, when the embodied agent guiding overlay the painting (IF and IT interfaces), Although, a few participants agreed that the virtual guide obstructed the area of explanation, most participants disagreed to it. Therefore this evidence supports some part of RQ 1, with the hypothesis that users do not mind if the embodied agent's position overlaps the painting.

Regarding the average time used by each interface, both teleport interfaces (IT and OT) have a lower average time used. Therefore, we accept RQ 2 that the teleport type moving interface helped reduce guidance time. However, these types of interfaces got many negative subjective ratings. Based on the interview, some participants mentioned that when the guide disappears, they didn't know where the guide would reappear. Also, they had to find the virtual guide position again most of the time when using the teleport type interfaces because there are no hints as to where the new position of the embodied agent would appear. The result from the questionnaires also confirmed that the teleport interfaces are not proper in position nor help the user find the area in context, and they do not look natural. We suggest that although the virtual guide is able to teleport in the real world, this is not a preferred interface for most of the participants. Therefore we should avoid using it or create a moving path or destination information for the agent so that the user will not lose the guide's position as the guidance progresses.

Although IF interface was rated that help the participant concentrate on both content description and area of explanation, but scores from the test results indicated that no guiding interface outperformed any other with respect to memory retention. Therefore, we reject RQ 3 that the different type of position and movement interface of the virtual guide does not affect memory recall of the image-related content. Based on the interview, many participants told us that there are too many items introduced so they can not remember them all. In many time they only remember either the image or the phase. Also, some participants said that they just enjoy the guide system but did not interested in the painting, so they did not try to remember the contents. Some participants mentioned that they did not like the outside position interfaces because they had to trace the virtual guide pointing to the image, while the inside position interfaces made it easier to find the part of the image because the virtual guide and the target position were in the same view.

By observing participants movement during the experiment, we found that participants who preferred IF and IT interfaces mostly stood around their starting position. There was only one participant who preferred IF interface that walk around the designated area. Six of eleven of participants who preferred OF interface and one from two of OT interface walked around the area, while half of the remaining participants stood around the start position. This indicates that if the virtual guide moves for a small distance (e.g. inside the painting area) the user will not move much from his/her standing position. However, if the virtual agent moves for a certain distance (e.g. from the left to right side), the user tends to follow that movement.

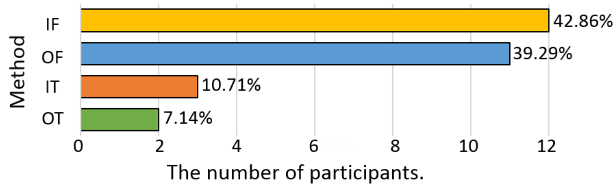

Figure 8: Plot showing the number of participants that ranked a particular method as the most preferred including Outside-Flying (OF), Outside-Teleport (OT), Inside-Flying (IF), and Inside-Teleport (IT).

5. Conclusion

In this paper, we proposed an AR guidance system for multi-section visual information using an embodied agent. We also implemented two positional interfaces and two moving interfaces for our virtual guide. We developed the system that tracks the location of a large piece of art and locates the position of regions interest, after which we provide guidance to the corresponding location. Our system included an embodied agent that moves and points toward the position, describes where and at what to look, and explains information for a specific part of the image.

Base on results from a user study, most of the participants preferred the Inside-Flying type, with the Outside-Flying pulling a close second since the flying interface was easier to track the position of a virtual guide than the teleport interface. Many participants were not concerned that the virtual guide obstructed the art when it stays in front of the painting as long as it does not occlude the part at which the user is looking. We found that the virtual guide that guides by moving inside the area of the painting helped participants find the position in the description immediately without having to trace the laser from the virtual guide's hand to the area in the description. This inside positioning interface also helped the user concentrate on a particular area of the painting. Regarding outside positioning interface, movement to another far positions of the virtual guide affected the user's movement.

In future studies, we will explore guiding interfaces for areas with obstacles such as other visitors to provide a dynamic guide system that automatically avoids or adapts its position based on its location and the order of descriptions.

Acknowledgements

We would like to thank all of the experiment participants for their time and participation. This research was funded in part by the Office of Naval Research Global, grant #N62909-18-1-2036.

References

[AI09] AI I.: Aitalk®, 2009. Retrieved May 20, 2019 from https://www.ai-j.jp/english. 3

[AKYT00] ANABUKI M., KAKUTA H., YAMAMOTO H., TAMURA H.: Welbo: An embodied conversational agent living in mixed reality space. In *CHI'00 extended abstracts on Human factors in computing systems* (2000), ACM, pp. 10–11. 1, 2

[AM11] AMBARISH MITRA OMAR TAYEB S. S. J. B.: Blippar : Augmented reality (ar) computer vision company, 2011. Retrieved November 28, 2018 from https://www.blippar.com/. 1

[AS07] ALLEN I. E., SEAMAN C. A.: Likert scales and data analyses. *Quality progress 40*, 7 (2007), 64–65. 5

[BAE*15] BICKMORE T., ASADI R., EHYAEI A., FELL H., HENAULT L., INTILLE S., QUINTILIANI L., SHAMEKHI A., TRINH H., WAITE K., ET AL.: Context-awareness in a persistent hospital companion agent. In *International Conference on Intelligent Virtual Agents* (2015), Springer, pp. 332–342. 2

[BDVW03] BEUN R.-J., DE VOS E., WITTEMAN C.: Embodied conversational agents: effects on memory performance and anthropomorphisation. In *International Workshop on Intelligent Virtual Agents* (2003), Springer, pp. 315–319. 2

[BPY08] BICKMORE T., PFEIFER L., YIN L.: The role of gesture in document explanation by embodied conversational agents. *International Journal of Semantic Computing 2*, 01 (2008), 47–70. 2

[BS04] BARAKONYI I., SCHMALSTIEG D.: *Ar puppet: Animated agents in augmented reality*. na, 2004. 1

[Cor16] CORPORATION M.: Microsoft hololens | the leader in mixed reality technology, 2016. Retrieved May 20, 2019 from https://www.microsoft.com/en-us/hololens. 2

[CSHO14] CAMPBELL A. G., STAFFORD J. W., HOLZ T., OÂĂŹHARE G. M.: Why, when and how to use augmented reality agents (auras). *Virtual Reality 18*, 2 (2014), 139–159. 1, 2

[Efr69] EFRON B.: Student's t-test under symmetry conditions. *Journal of the American Statistical Association 64*, 328 (1969), 1278–1302. 5

[Gir92] GIRDEN E. R.: *ANOVA: Repeated measures*. No. 84. Sage, 1992. 5

[HBB*68] HALL E. T., BIRDWHISTELL R. L., BOCK B., BOHANNAN P., DIEBOLD JR A. R., DURBIN M., EDMONSON M. S., FISCHER J., HYMES D., KIMBALL S. T., ET AL.: Proxemics [and comments and replies]. *Current anthropology 9*, 2/3 (1968), 83–108. 2

[Hol79] HOLM S.: A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* (1979), 65–70. 5

[IOI*01] ISHIGURO H., ONO T., IMAI M., MAEDA T., KANDA T., NAKATSU R.: Robovie: an interactive humanoid robot. *Industrial robot: An international journal 28*, 6 (2001), 498–504. 3

[Jap14] JAPAN U. T.: Unity-chan!, 2014. Retrieved May 20, 2019 from unity-chan.com/. 2

[KGKW05] KOPP S., GESELLENSETTER L., KRÄMER N. C., WACHSMUTH I.: A conversational agent as museum guide–design and evaluation of a real-world application. In *Intelligent virtual agents* (2005), Springer, pp. 329–343. 2

[Mas04] MASSARO D. W.: Symbiotic value of an embodied agent in language learning. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on* (2004), IEEE, pp. 10–pp. 2

[Mas16] MASON M.: The mit museum glassware prototype: Visitor experience exploration for designing smart glasses. *Journal on Computing and Cultural Heritage (JOCCH) 9*, 3 (2016), 12. 2

[MGR14] MACEDONIA M., GROHER I., ROITHMAYR F.: Intelligent virtual agents as language trainers facilitate multilingualism. *Frontiers in Psychology 5* (2014), 295. 2

[MLF09] MAARTEN LENS-FITZGERALD CLAIRE BOONSTRA R. V. D. K.: Layar : Augmented reality | interactive print, 2009. Retrieved November 28, 2018 from https://www.layar.com/. 1

[MMT*08] MIYASHITA T., MEIER P., TACHIKAWA T., ORLIC S., EBLE T., SCHOLZ V., GAPEL A., GERL O., ARNAUDOV S., LIEBERKNECHT S.: An augmented reality museum guide. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality* (2008), IEEE Computer Society, pp. 103–106. 2

[NZB00] NOMA T., ZHAO L., BADLER N. I.: Design of a virtual human presenter. *IEEE Computer Graphics and Applications 20*, 4 (2000), 79–85. 2

[RSTB15] RING L., SHI L., TOTZKE K., BICKMORE T.: Social support agents for older adults: longitudinal affective computing in the home. *Journal on Multimodal User Interfaces 9*, 1 (2015), 79–88. 2

[SFT96] SHELDON M. R., FILLYAW M. J., THOMPSON W. D.: The use and interpretation of the friedman test in the analysis of ordinal-scale data in repeated measures designs. *Physiotherapy Research International 1*, 4 (1996), 221–228. 5

[STA*10] SWARTOUT W., TRAUM D., ARTSTEIN R., NOREN D., DEBEVEC P., BRONNENKANT K., WILLIAMS J., LEUSKI A., NARAYANAN S., PIEPOL D., ET AL.: Ada and grace: Toward realistic and engaging virtual museum guides. In *International Conference on Intelligent Virtual Agents* (2010), Springer, pp. 286–300. 2

[TRO*19] TECHASARNTIKUL N., RATSAMEE P., ORLOSKY J., MASHITA T., URANISHI Y., KIYOKAWA K., TAKEMURA H.: Evaluation of pointing interfaces with an ar agent for multi-section information guidance. In *Proceedings of IEEE Virtual Reality and 3D User Interfaces* (2019). 3

[VKL16] VAINSTEIN N., KUFLIK T., LANIR J.: Towards using mobile, head-worn displays in cultural heritage: user requirements and a research agenda. In *Proceedings of the 21st international conference on intelligent user interfaces* (2016), ACM, pp. 327–331. 2

[VLP*03] VACCHETTI L., LEPETIT V., PAPAGIANNAKIS G., PONDER M., FUA P., THALMANN D., MAGNENAT-THALMANN N.: Stable real-time interaction between virtual humans and real scenes. In *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on* (2003), IEEE, pp. 449–456. 2

[Wik09] WIKIPEDIA: Netherlandish proverbs, 2009. Retrieved May 20, 2019 from https://en.wikipedia.org/wiki/Netherlandish_Proverbs. 3

[Woo07] WOOLSON R.: Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials* (2007), 1–3. 5