

Experiences Employing Novice Wizard Operators in a Gallery Setting

Steven Dow and Blair MacIntyre

School of Interactive Computing and GVU Center
85 TSRB, 5th St, NW, Atlanta, GA 30308
{steven,blair}@cc.gatech.edu

Abstract. During a 11-week installation of an immersive augmented reality (AR) experience at an art and technology gallery, we faced the practical challenges of preparing for live public audiences and of training and motivating a group of nine modestly-paid undergraduate museum docents to not only usher the show, but to perform as “wizards”. The docents played an integral part of the AR Façade experience, replacing imperfect recognition technology to achieve speech and gesture interaction for the players. Our investigation of the wizard docents revealed insights into two conceptually different wizard-of-oz implementations. The docents initially resisted the interface option that required more cognitive load, but after learning how to make it work, many preferred that interface because it provided them more agency within the system and freedom to diversify the experience for audiences.

Keywords: Immersive entertainment, gallery installations, user experience, wizard-of-oz methods, augmented reality.

1 Introduction

Interactive entertainment experiences—especially in amusement parks or gallery settings—often employ full-time operators to usher visitors in and out of the experiences [5][7]. More recently, interactive media have looked to trained actors performing behind the scenes in real-time as animated characters or robots, such as [8] and [9], very similar to the hidden wizard in the movie *The Wizard of Oz* [1]. Other experiences use actors more overtly, such as *Can You See Me Now?* where online players try to locate physical actors who are moving around a city [3]. In this paper we report on our experience creating an intermediate role for “wizard” docents, who hold responsibilities somewhere between an usher and an actor.

Our discussion centers on an 11-week public gallery installation of AR Façade, a first-person immersive augmented reality (AR) experience where a player enters a physical apartment with a head-mounted display and interacts using speech and gesture with two fictitious characters [4]. Rather than use speech/gesture recognition technology—which can be problematic, particularly with diverse users in non-controlled settings—we tasked our wizard docents with accomplishing this interaction.

In this paper, we describe the modifications we made to the AR Façade experience to prepare for a long-term deployment at the Beall Center for Art and Technology at the University of California Irvine [2]. The gallery installation would be held to higher standards than our initial lab prototype and it would need to entertain audiences of people, not just the individuals who would participate in the immersive aspect. More importantly, the show would have to be easy to operate by non-technical, undergraduate, gallery docents who were not part of the research team, nor particularly motivated. We set out to investigate whether and how docents would fulfill this role by loosening the bounds of the wizards’ task and establishing two conceptually different wizard interfaces. We share data about how the nine docents learned to use and adopt the two interfaces over the course of the show. We found

that the interface that required more cognitive processing actually encouraged deeper involvement in the story and in reading the tone of the players and audience members.

2 The Gallery Installation of AR Façade

In this section, we describe our design choices for the three-month Beall Center deployment of AR Façade. We believe our experience of upgrading the prototype into a real deployment can be useful for producers and designers of immersive experiences. We built the immersive interface to be robust and to withstand 11-weeks of public use by audiences of all shapes and sizes and ages. We designed the physical layout so that audiences and the wizards would get a strong sense of what the player experienced. Finally, we experimented with the wizard infrastructure, providing two methods of asserting player inputs so that we could investigate wizard behavior.

2.1 Physical Setup

AR Façade would be part of a free-to-the-public art gallery for 11 weeks, so we devoted a fair amount of time to the construction of the head-mounted display (HMD) (Fig. 1 *left*) and the physical layout of the space (fig. 1 *middle*). We constructed the head-mounted display from third-party components and custom-created mounts to be robust enough to withstand mishandling. We tracked the HMD using the IS-1200 Vistracker and large visual markers mounted on the 16' tall ceiling. Florescent lights lined the perimeter of the apartment at 10' high and provided an even distribution of light across the ceiling and enough illumination for the stage. We did not encounter any noticeable tracking problems—such as jitter and temporary loss of tracking data as we did with our initial prototype—despite losing a few markers from the ceiling.



Fig. 1. (left) The head-worn display. (middle) Audience's view into the physical apartment with monitor showing the player's perspective. (right) Player's HMD view shows the Trip character standing behind the physical bar. Note that the black scrim blocks the view of objects beyond the wall.

We focused on the physical layout of the space, because we wanted the experience to be visually appealing and enjoyable to groups as well as individuals. We used black scrim along one of the long walls of the apartment behind the bar. Since it was light inside the apartment and dark outside, audiences could stand outside the wall and easily see the player's activity, while the player would not be able to see outside though the wall (especially when viewing through the HMD) (fig. 1 *right*). Moreover, the black walls and the black-painted bar helped maintain the occlusion effect when Trip walks behind the bar; players know the bar is there, but the top of the bar is difficult to distinguish from the wall.

We used a wireless video transmitter to transmit the player's HMD view to a TV monitor sitting to the right of the stage. The wizard—sitting just outside the door of the

apartment—could also see the same HMD view, along with a second monitor with video from a ceiling mounted camera. The audio turned out to be one of the biggest challenges. We were competing with several other loud exhibits in the space, so we wanted the audience (and wizard) to be able to hear the dialogue between the player and the characters. To make this work, we used two audio transmitters (one to amplify the player’s voice and one to transmit the backpack’s computer audio) and a mixing board to mix the two sources and disseminate it to speakers for the audience and headphones for the wizard.

2.2 Wizard Infrastructure

We made a number of changes aimed at supporting the undergraduate docents who would be in charge of operating and wizarding the experience. Most of the enhancements were practical (providing a start-up script, building in network checks, etc.), but we also experimented with the underlying infrastructure of Façade. Dissecting the system architecture of the original Façade (described here [6]), we thought it would be interesting to tear out the natural language parser (NLP) and build a wizard interface that could be used to directly trigger the higher-level constructs (discourse acts). This approach was plausible because Façade’s AI engine does not explicitly map user inputs to specific statements and actions by the two virtual characters. Instead, it models the characters’ emotional states and attempts to choose lines of dialogue based on local and global contexts. There are about 30 possible discourses (e.g. flirt, agree with, etc.) with optional parameters that can be expressed as local context by the player.

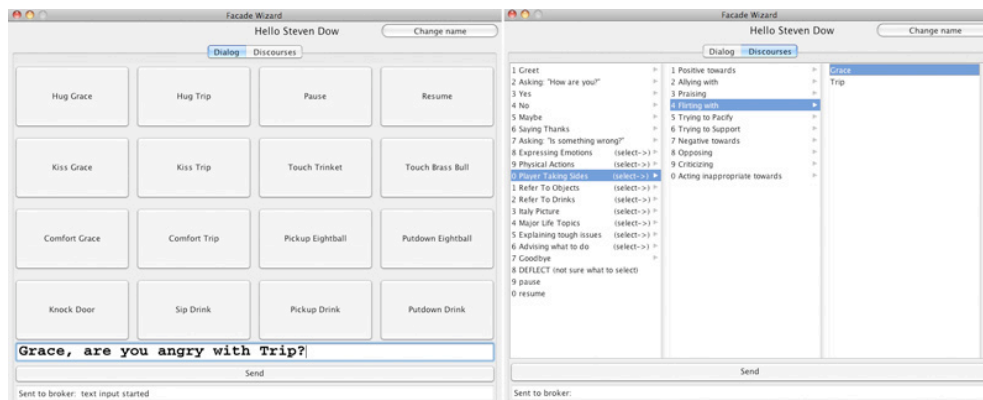


Fig. 2. (left) Dialogue wizard interface provides a text entry field and buttons for indicating player specific gestures. (right) Discourse wizard interface provides a hierarchy of discourses or player intentions, such as “Player Taking Sides->Flirting with->Grace”.

The new wizard interface had two panes—“Dialogue” and “Discourses”—both available to the wizard at any time. The Dialogue pane included a text entry field for typing in what players said and large buttons for indicating specific player actions (Fig. 2 left). This mirrored our original wizard interface, including the visual and audio feedback to wizards indicating the maximum number of chars that could be entered at one time.

The Discourses interface provided a hierarchical listing of the higher-level discourses (Fig. 2 right); rather than typing out what the player says, the wizard selects something that matches what the player *means* (e.g. “Player takes sides by flirting with Grace”). The Discourses interface went through a number of iterations to improve its usability, including incorporating feedback from the docent wizards after the first couple weeks of the show. The

current interface represents the discourse hierarchy across three columns that can be navigated through any combination of number keys, arrow keys, and the mouse.

3 An Investigation of the Wizard Docents

We initially created the Discourses interface because we believed we could improve on the fundamental time delay between a player speech utterance and the responses by the characters. We wanted to know if wizards could use the Discourse method, and how it compared to the typing interface. Would it be as fast, slower or faster than the other method? How long would it take to learn the discourses? And, how would it affect the experience for players and audiences?

Throughout the 11-week installation we recorded wizard activity and conducted open-ended interviews with the nine docents, non-technical female art students between the age of 19 and 22. After a short preliminary interview, we trained each wizard to use both versions of the wizard interface and instructed them to create the player experience however they saw fit. During the first 9 weeks of the show—the period where we were not onsite—we allowed the docents to use either wizard interface. During the final 2 weeks, we noticed that most wizards relied on the Dialogue interface, so we forced the docents to learn and use the Discourse interface. We conducted numerous open-ended interviews during the final 2-week period. In this section share some of the insights gained from these interviews, supported by the usage patterns and qualitative sentiments expressed by the docent wizards.

3.1 Wizard Interface Usage

Examining the game episodes for AR Façade for the 11-week period we can visualize the activity of the wizard docents. There were 106 full episodes—some repeat players—distributed across the nine wizards. The wizards were free to use either the Dialogue or Discourse interface, and the data shows the former was used more frequently, although W8 used the Discourse interface nearly half the time (Fig. 3). Across all episodes and all wizards, lines of dialog were entered 2.9 times per minute while discourses were entered 0.5 times per minute on average.

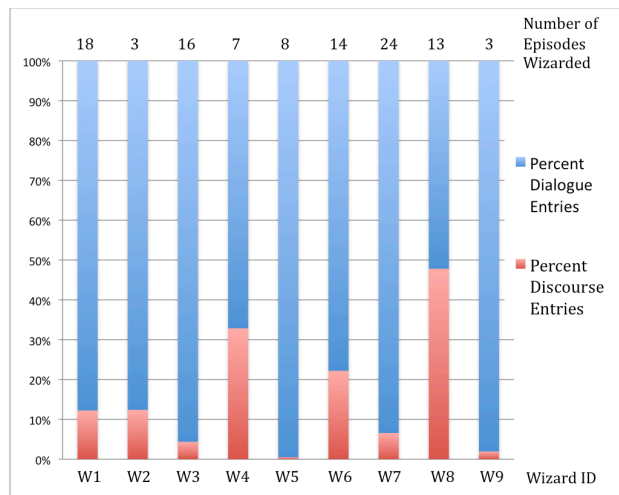


Fig. 3. Wizard activity during 11-week installation of AR Façade across nine wizard docents.

3.2 Wizard Insights on the Two Interfaces

From the wizard interviews, we understood why the wizards generally utilized the Dialogue more than the Discourse interface. There was a consensus among wizards that the Dialogue interface was simply easier; with the Discourse interface “picking stuff out requires more thinking” (W3). According to W4, “it forces your mind to kind of think in a different way, of not just directly translating specifically what they’re saying but kind of attributing it to a larger category of emotion or actions. It depends a little more on your interpretation.” And W6 explains, “I feel like I have to interpret more what the players are saying. So it’s like more involved, like – you have to pay more attention.” W8 pointed to one reason to use the Discourse interface: if there was an audio problem, she didn’t have to hear the exact phrasing to pick up on the player’s intentions.

So the Discourse category selection appears to require more cognitive load, but does it reduce the time delay issue and help the conversation flow better? According to W1, “typing it out takes a little bit longer than searching and clicking,” while W4 says “it takes a little longer to sort through them and find the right one then it does to just immediately translate what they’re saying into text.” Answering this question will require a more detailed conversation analysis, but based on our initial data analysis, we suspect that Discourse selection is faster for more verbose statements and slower for shorter statements.

On the other hand, some wizards reported that the Discourse interface gave them more control over the characters and the course of events. W9 claimed “you could definitely shape the player’s experience”. W1 even went further, almost describing the characters as puppets: “I’ll make Trip talk about the picture again and hopefully he will kind of guide [the player] over.” W6 said it was her role “to translate what the people are saying so that Trip and Grace can understand it”, anthropomorphizing the virtual characters.

3.3 Wizard Interactions with Players and Audiences

The interviews not only revealed the affordances of each interface, they revealed how the wizards appropriated the system, especially after we enforced the use of the discourse interface. While the Dialogue interface required constant typing and emulation of player speech—especially since the words appear on the player’s HMD—the Discourse interface afforded a degree of experimentation. W8 said if there is a “lull in the conversation... I will select something like Therapy, just to offer a little variation. ... because some people would be a little passive in their interactions.” For the most part, the wizards were trying to help, but sometimes they got bored and clicked discourses just to spice things up: “I clicked ‘have sex’ or something because I was hoping that some big explosive thing would happen... I thought that would be fun to see, because Grace seems like kind of an intense chick” (W9).

There were some consequences for this experimentation, as W4 found when she tried using the ‘oppose Trip’ discourse: “I guess that was too strong of an emotion, because [Trip] kicked him out. haha!” Wizards reported that the system did not always respond as expected, that they wanted more nuance, and that sometimes it felt their actions did not have any affect.

In their role as wizards, the docents learned to become very perceptive of the state of the player. “I can tell when someone feels awkward or when people are getting really annoyed by just like the tone of their voice” (W3). Some wizards were so attuned to how players felt, the emotions were palpable, “I could hear [the player’s] breathing pattern and ... I could almost feel her getting uncomfortable and it’s kind of bizarre because you can hear the audio and then you can see what they’re seeing and I almost started to get uncomfortable!” (W9).

The wizards had different theories on what would be most fun for players and audiences. W7 said she would tell players ahead time “the more you mess with [Trip and Grace] the more entertaining it will be.” W1 said she also encourages players to play with it “because a lot of the times people tend to like stay in the middle of the room or they don’t get

near the objects. ...that way there's a little bit more for me to do in the back too." W3 would flatly reveal her role, even invite players and audiences to see behind the curtain. "I would tell people that I'm gonna be back there.... so that they can get the reaction that they want." (W3)

Some wizards felt it was important to preserve the illusion, as W8 states "I would say the illusion is necessary because it frees up this whole possibility of what could happen if they like slapped Trip or something." According to W6, if players saw the wizards they "seemed kind of disappointed. It's like magic tricks. When you find out it's not fun." (W6)

4 Discussion

While the question of whether the Discourse interface reduced input delay remains to be answered in detail, we learned a great deal by exploring alternative wizard infrastructures. As expected, the Dialogue typing interface did not require as much mental attention, so the wizards initially preferred it. Once we forced them to use the Discourse interface and they starting experimenting and understanding how it worked, the wizards became more attuned to the course of the story and the emotional level of the players and audience members. While the Dialogue interface provided a good design from one standpoint (easier to learn and simpler to use), the Discourse interface engaged the wizards as performers of sorts, enabling them to guide the direction of the experience as they saw fit.

5 Conclusion

In our investigation of the nine non-technical wizard docents that ran AR Façade for 11 weeks, we found that (with training) the wizards became receptive to the Discourse interface where they choose categories of intentionality, not just type in player statements. Although it was more difficult to master, the discourse method gave wizards more control over the course of events. Our wizards reports they would deviate from and actively assert players intentions to create a more appealing experience for visitors, as well as for their own enjoyment.

References

1. Baum, F. The wonderful wizard of oz. George M. Hill, Chicago (1900)
2. Beall Center for Art and Technology, <http://beallcenter.uci.edu/>
3. Benford, S., Crabtree, A., Flinham, M., Drozd, A., Anastasi, R., Paxton, M., Tandavanitj, N., Adams, M., and Row-Farr, J. Can you see me now?. ACM Transactions on Computer-Human Interaction. 13, 1, 100-133 (Mar. 2006)
4. Dow, S., Mehta, M., Lausier, A., MacIntyre, B., and Mateas, M. Initial Lessons From AR Façade, an Interactive Augmented Reality Drama. In ACE '06: Proceedings of the 2006 ACM SIGCHI Int.l conf. on Advances in Computer Entertainment, ACM Press, (2006a)
5. Koleva, B., Taylor, I., Benford, S., Fraser, M., Greenhalgh, C., Schnadelbach, H., vom Lehn, D., Heath, C., Row-Farr, J., and Adams, M. Orchestrating a Mixed Reality Performance. In CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 38-45, New York, NY, USA. ACM Press. (2001)
6. Mateas, M. and Stern, A. Façade: An experiment in building a fully-realized interactive drama. In Game Developer's Conference: Game Design Track. (2003)
7. Pausch, R., Snoddy, J., Taylor, R., Watson, S., and Haseltine, E. Disney's Aladdin: First Steps Toward Storytelling in Virtual Reality. In SIGGRAPH '96. ACM Press. (1996)
8. Quasi the Robot, http://www.etc.cmu.edu/projects/ibi/platform_hardware.htm
9. Turtle Talk with Crush, http://en.wikipedia.org/wiki/Turtle_Talk_with_Crush