

Understanding Engagement: A Mixed-Method Approach to Observing Game Play

Steven Dow

School of Interactive Computing
College of Computing

Georgia Institute of Technology

85 5th Street, NW, Atlanta, GA 30332

steven@cc.gatech.edu

Blair MacIntyre

School of Interactive Computing
College of Computing

Georgia Institute of Technology

85 5th Street, NW, Atlanta, GA 30332

blair@cc.gatech.edu

Michael Mateas

Computer Science Department
University of California, Santa Cruz

1156 High Street, M/S SOE3

Santa Cruz, CA 95064

michaelm@cs.ucsc.edu

ABSTRACT

We want to believe HCI has much to offer the game industry, but it goes beyond simply acknowledging differences between the HCI notion of a user and the game design notion of a player. People engage in play in very different ways—and so our evaluation methods should reflect this diversity. Game designers and researchers must be careful in establishing common frameworks for evaluation that stymie creativity and perpetuate standard game conventions. Due to the situated nature of games, we argue for mixed method data gathering and ethnographic-style analyses to explain players' interpretations and behaviors. From this evaluative stance, games also prevail as excellent tools for the study of human experience. To support our position, we share our reflections on our empirical evaluation of the critically acclaimed interactive drama *Façade* [11] and our augmented reality (AR) adaptation of *Façade* [4]. The *Façade* “game” actually serves as a research tool for our investigation of immersive interfaces and the deeper human issue of embodiment.

Categories and Subject Descriptors

H.5.2 [User Interfaces]: *Evaluation/methodology*.

General Terms

Design, Experimentation, Theory

Keywords

Games, ethnography, mixed-methods, evaluation, interactive drama, immersive interfaces, augmented reality.

1. INTRODUCTION

While the concept of marrying the fields of HCI and computer games has been a topic of much discussion [7, 15], we must be wary how to proceed. If evaluation techniques for games go the direction of traditional HCI, it could lead to stale variations of typical game conventions. Likewise, game designers should not be so quick to dismiss the value of observation offered by skilled HCI experts and ethnographers.

At the forefront of this intriguing amalgamation of disciplines is the games user research lab at Microsoft, whose analysis of 3000+ hours of *Halo 3* gameplay by 600 gamers led to significant improvements in everything from simple bug fixes, to new subtle design features to prevent players from running out of ammo, to adjustments in enemy power to maintain the right level of challenge [14]. The usability testing for *Halo 3* identified

problems by recording as much any in-game data as possible so that oddities could be visualized and understood. Game testing is necessary and valuable, but inevitably it comes at late stages of development, obviating the possibility for major design shifts.

Lazzaro and her team have been using video analysis techniques and interview methods towards understanding the range of player emotions and reasons that people play games [8]. Again Lazzaro's studies focus on finished (or nearly finished) games, perhaps too late to influence the emotional impact of that particular game, but a step in right direction of calling for more varied play to account for differences in skill level, emotional state, and motivation.

So perhaps we should first consider the “why?”, “how?”, and “when?” behind the premise of evaluating games. Are we actually looking to make innovations in games or are we merely affirming indoctrinated assumptions in a game's design? What information is interesting to collect and what would it say the about game? Beyond simple usability issues (such as realizing players cannot find ammo in a certain level [14]), what kind of feedback is usefully integrated back into the game's design and when is it most useful? How do evaluations break away from underlying assumptions built into game designs so that revolutionary changes can happen? In this paper, we strongly advocate ethnographic-style approaches (in combination with in-game quantitative measures) because it yields rich information at all levels—from conceptual and mechanical, to purely interface design issues. For commercial games, observational methods should come early enough to allow for significant design changes; but as we share below, these approaches are valuable for game researchers hoping to learn more about human-centered research topics.

Our experience with game evaluation comes from our investigations of AR *Façade*, an embodied version of the interactive drama *Façade* [11], authored by Mateas and Stern. We exploit *Façade*—a game experience based largely on social interaction with two characters, Grace and Trip—to conduct a comparative evaluation of standard desktop interaction with immersive “augmented reality” (AR) interaction [4]. AR *Façade* pushes technological frontiers in artificial intelligence and physically embodied interaction while we also use it as a research *instrument* in the study of narratology [10], ludology [10], play [12], funology [1], and embodied interaction [3]. And while our experiences may not be typical of game evaluation in industry, we are working with a “real”, complete game environment so we can assert that our findings are relevant. However, like most game user studies, our investigations come after the public release of *Façade*, so the results did not explicitly impact the underlying game design for *Façade*. It remains an open-challenge to quickly

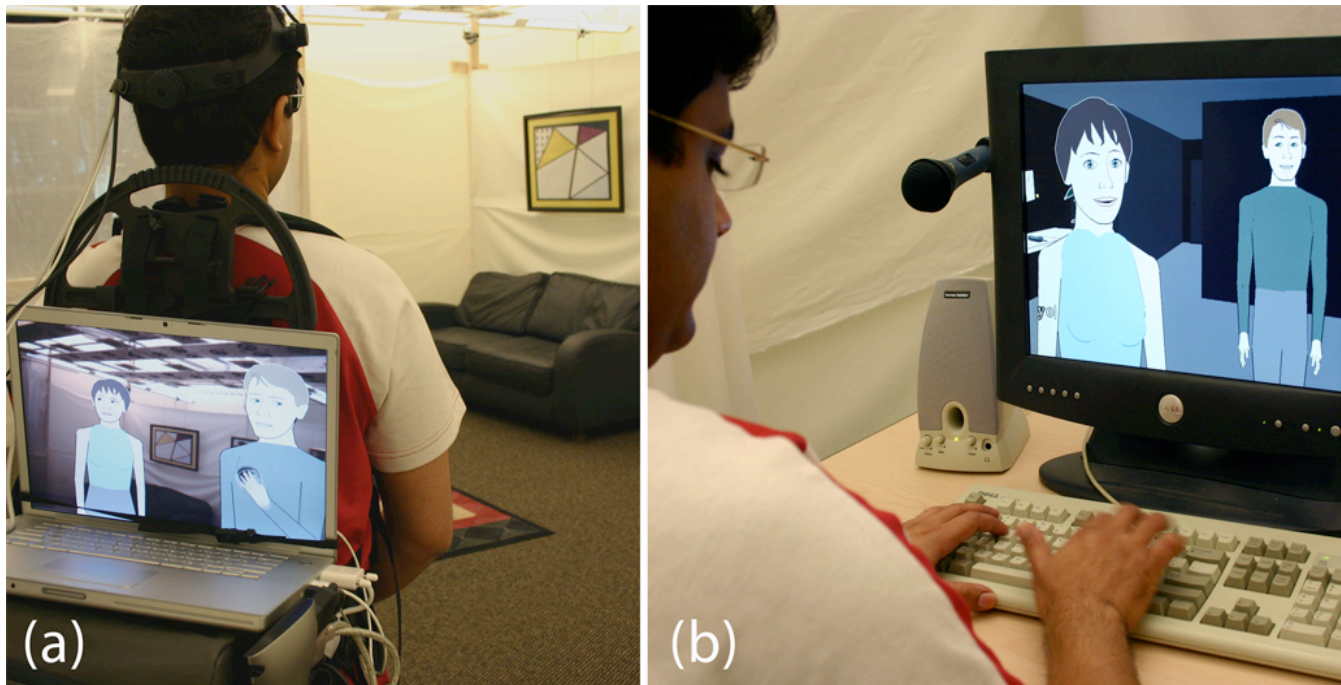


Figure 1: a) Augmented Reality (AR) Façade: The player wears a head-mounted-display and experiences the characters (Trip and Grace) in a physical apartment, interacting through voice and gestures. b) Original desktop Façade: The player types out conversation, navigates with keyboard arrows, and touches using mouse interaction.

prototype games so that useful feedback can be obtained when it matters most.

Our position paper has two arguments. We argue that games can be invaluable research instruments—Petri dishes in a sense—for the study of human experience. Second, we claim that games are phenomenological experiences that are best understood and analyzed through an ethnographic lens—untainted by preconceived game design constructs. We support our arguments by sharing our work on AR Façade and our study of player engagement and presence.

2. GAMES AS RESEARCH TOOLS

Similar to Lazzaro's studies of emotion [8] in a variety of popular games and Lee et al's studies of social interaction [16] in *Second Life*, we find games to be convenient tools for human investigation. Rather than asking usability type questions, we approached the Façade experience with a specific research agenda. We were interested in the impact of immersive interfaces on game experience and Façade provided an excellent opportunity to get traction into questions of presence and engagement. We developed an immersive AR version, complete with a physical apartment to match the virtual desktop apartment of Grace and Trip (see Figure 1). We designed a qualitative study to compare desktop interaction with the AR version and the results were quite illuminating. Our study found that while the more immersive AR interface increased most players' sense of "presence" over desktop interaction, this heightened presence did not always lead to increased engagement and fun.

Incidentally, we learned as much about how players enjoy the genre of interactive drama as we did in the differences between

interfaces. If we happened to be in the midst of the design process for Façade, we could make serious improvements to the navigation, the natural language parser, and other constructs in the game. By regarding Façade as a research tool, we were able to look beyond just this game and draw out philosophical insights.

3. GAMES AS SITUATED EXPERIENCE

All games inculcate values and conventions whether the designers realize it or not. Game designers—like all software designers—build in basic assumptions about the navigation mechanics and world values [2, 5]. Players approach games with a life's worth of experiential "baggage"—memories of other games and other media, cultural upbringings, proficiency with physical game controllers, etc.—which influence how play unfolds. Video games present phenomenological "encounters" that are played out according to personal experience and situational context. A great game manages to build on cultural and social "capital" and physical "affordances" so that players can engage at different levels of the psyche.

Unlike traditional software where HCI evaluative heuristics can be used to determine if the user accomplishes a particular task, games must consider the context of the experience. "Checklist" evaluations do not do justice to understand the emotional depth of game experiences. Rather, open-ended data gathering and interpretive methods provide a deeper glimpse of what might be happening and can result in more meaningful feedback. Like any experience in life, video games provide an opportunity for interpretation, decision-making, and action—it should fall upon an ethnographer or HCI expert to uncover subtle behaviors and hidden meanings when a player encounters a game. In our study of Façade, to avoid preconceived categories of results, the

evaluator (the first author of the paper) was familiar with Grounded Theory [13] analysis and was also disconnected from the game’s authorship and design.

4. THE AR FAÇADE EXPERIENCE

To provide a little more background on Façade, we will briefly describe the drama and interaction in this first foray into “interactive” drama, but much more can be found in other papers [4, 11]. The drama revolves around the collapsing marriage of the protagonists Grace and Trip. Seemingly invited over for a friendly evening, the player finds herself in the middle of an argument and is forced by Grace and Trip into playing psychological “head games” with them. The player, choosing her own name and gender, may react to the experience with hilarity or anger, or play a number of roles from councilor to devil’s advocate. Unlike most games, the players are not given a clear goal; the player invents goals for herself as the interaction with the characters unfolds.

The original version of Façade was designed for desktop computing, so the player uses arrow keys to move around the virtual apartment, the mouse to “touch” things in the space, and the normal keyboard input to communicate to Grace and Trip (see Figure 1(b)). In the AR adaptation of Façade, the player uses speech to communicate, but navigation occurs in an actual physical apartment built to match the virtual desktop apartment. In AR, the entire world is physical except for Trip and Grace, who are aligned with the physical space and superimposed on the world using a video-see-through head-mounted display (HMD) worn by the player (see Figure 1(a)). The player can enact game actions such as hugs and kisses, by performing a physical gesture. Both the speech and gesture recognition are accomplished using a hidden operator behind the scenes, also known as the Wizard-of-Oz method, so that we could get traction into our research questions without getting bogged down with too many with interface imperfections.

We can now go into further detail about our “mixed-method” data gathering methods and our “grounded” analysis of the different versions of Façade. We also share our findings and reflect on the impact of our methods on the results.

4.1 “Mixed-Method” Data Gathering

Façade is specifically designed to be replayable and indeed the storylines and character dialog never quite happen the same way twice. We used this fact to our advantage and asked twelve participants to play each version of the interface for comparison. They actually played three different versions; we added a speech-based desktop version to tease out the difference between walking and speaking from sitting and typing. After each game episode, the researcher and the player jointly watched key moments on captured video. During these open-ended *retrospective* interviews, the players were prompted with statements like “So, tell me what happened here.” The researcher tried to avoid asking leading questions that could influence the player’s thinking. All the interviews were recorded on video so they could be analyzed at a later time. In addition to the interviews, we collected a variety of quantitative data, including position/rotation points from the characters and players, conversation logs, and Likert-scale survey questions. We understood that survey questions inevitably reflect the biases of the investigators, but we felt it was acceptable to gather this additional data at the very end of the 3-hour session.

4.2 “Grounded” Analysis

Although one might question how we can compare two fairly distinctive game episodes, our open ethnographic-style interviews and observations revealed differences during analysis. We analyzed over 36 hours of game play and interviews using a ground theory approach [13]. Everything was transcribed and all statements were labeled with notes. The notes were printed onto tiny pieces of paper and spread out so that a team of researchers could make groupings. Eventually, some interesting themes arose and we could back them up with specific quotes and anecdotes from the recorded episodes. The grounded theory approach is indeed subjective, but a trained ethnographer will be able to make observations that system designers are not able to see. The key is to avoid making any presumptions about what should happen; the process should reveal what happens and where players are experiencing emotion—whether it’s fun, frustration, excitement, wonderment, or boredom. Even with a relatively small number of participants (only 12 versus the 600 used in the *Halo 3* study) we were able to provide rich descriptions about specific actions and explicate the players’ interpretations.

4.3 Findings

As we stated up front, the purpose of our study was to compare desktop computing with a more embodied form of interaction. We found that most of the players felt a stronger sense of presence with the characters and the environment when they were able to use their bodies. Most intriguing, some players actually preferred the less immersive desktop interaction for this very reason—they were less connected and had more emotional distance from the role they were playing in the social scenario. While the immersive version was implemented well enough for people to enjoy, it did not automatically make the experience more engaging. This was particularly interesting in light of contrary assumptions made by some in the presence research community [9].

Our findings include qualitative observations on how players form styles of play and strategies for adapting to this new genre of game, and also quantitative differences in how closely players stand away from the two characters. The immersive version appears to induce players to stand closer to Trip and Grace, and this is explained to some extent by qualitative data. Players found it difficult to move around in AR, but easy to just turn their heads. This closer interpersonal distance in AR may have contributed to stronger sense of social presence and—because of the specific social context of an arguing couple—led some players to feelings of social awkwardness.

While these findings are fascinating from our research perspective, we believe our analysis could be valuable for a redesign of Façade. More likely, we will try to adopt these methods much earlier in the design and development for such an experience in the future. In particular, the genre of immersive interactive dramas can benefit from two domains of feedback: language patterns and how to structure interaction. We started to learn about the affordances of Façade’s dialogue—specific statements that would elicit a very common set of responses by players. For example, at one point in the experience the Grace character sometimes exclaims “No one ever asks me what I want.” This nearly always results in a player reaction—not in the exact the phrasing, but similar enough—along the lines of “What do you want Grace?” Grace’s reaction is silence. These common “dialogue breakdowns” are actually conceptually similar to not

picking up the spare ammo in a first-person shooter, although the methods of smoothing over the conversational flaws will be much different (and might eventually require expertise in Conversational Analysis [6]).

As far as how to structure interaction, desktop Façade was safely mediated using the well-understood interaction of the mouse and keyboard, while “augmented reality” Façade was an attempt to build on people’s knowledge of real-life interaction. Unsurprisingly, it fell short of expectations and players found themselves wondering how to effectively interact. They felt presence in the medium, but didn’t have proper “hooks” for knowing what could (and could not) be attempted. In other words, the edges were unclear.

Again, we point to the value of rich qualitative data gathering combined with quantitative logging and grounded analysis methods for understanding a game. Anecdotal evidence is much simpler to gather than statistical evidence and, for many people, more compelling. Moreover, the evaluative approach fits with both of our arguments for games; the “situated” nature of games is in fact why games are interesting as research tools.

5. CONCLUSIONS

There are two points we want to shine through in this position paper. The first is that games can be excellent research tools for learning about humans. We used Façade to explore different levels of immersion in human-computer interfaces, only to find that explicit mediation may be desired for players to engage in certain immersive experiences. Second, a game experience can be philosophized and analyzed as a contextual situation, as is any direct experience in life. Towards both points, a mixed-method approach revolving around ethnographic-style observation can be invaluable for player insight. While the quantitative data may reveal that things are happening, the qualitative insights show why they happen. HCI can offer game design an understanding of how players engage, independent of the intentions of the games designers. Similarly, computer games—unlike most software systems—are designed for users to do the unexpected, to deviate from the path, to search out the edges, to get satisfaction from a good challenge, and to potentially reach new emotional depths. Perhaps a marriage of HCI and game design is indeed possible.

6. ACKNOWLEDGMENTS

Our thanks to Andrew Stern for developing Façade (along with Dr. Mateas), and to Manish Mehta for helping build AR Façade.

7. REFERENCES

- [1] Blythe, M.A., Overbeeke, K., Monk, A.F. and Wright, P.C. *Funology: from usability to enjoyment. Kluwer Academic Publishers*. Dordrecht, Boston, London, 2003.
- [2] Bogost, I. *Persuasive Games: The Expressive Power of Video Games. Cambridge: MIT Press*, 2007.
- [3] Dourish, P. *Where the Action Is: The Foundations of Embodied Interaction. Cambridge: MIT Press*, 2001.
- [4] Dow, S., Mehta, M., Harmon, E., MacIntyre, B., and Mateas, M. Presence and Engagement in an Interactive Drama. In *Conference on Human Factors in Computing Systems (CHI'07)*, San Jose, CA, U.S.A, 2007.
- [5] Flanagan, M., Howe, D., and Nissenbaum, H.. Values at Play: Design Tradeoffs in Socially-Oriented Game Design. In *Conference on Human Factors in Computing Systems. CHI'05*, 2-7 April, Portland, Oregon, 2005.
- [6] Hutchby, I. and Wooffitt, R. *Conversation Analysis*. Polity Press, 1988.
- [7] Jørgensen, A. H. 2004. Marrying HCI/Usability and computer games: a preliminary look. In *Proceedings of the Third Nordic Conference on Human-Computer Interaction* (Tampere, Finland, October 23 - 27, 2004).
- [8] Lazzaro, N. (2004). Why We Play Games: Four keys to more emotion without story. *Online Article*, July 2006. http://www.xeodesign.com/xeodesign_whyweplaygames.pdf
- [9] Lombard, M. and Ditton, T., At the Heart of It All: the Concept of Presence. In *Journal of Computer-Mediated Communication*, 3(2), 1997.
- [10] Mateas, M. and Stern, A. Build it to Understand It: Ludology Meets Narratology in Game Design Space. In *Proceedings of the Digital Interactive Games Research Association Conference (DiGRA '05)*, Vancouver B.C., June, 2005
- [11] Mateas, M. and Stern, A., Façade: An Experiment in Building a Fully-Realized Interactive Drama, In *Game Developer's Conference: Game Design Track*, 2003.
- [12] Salen, K. and Zimmerman, E. *Rules of Play*. MIT Press, 2003.
- [13] Strauss A, Corbin J. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Sage, 1990.
- [14] Thompson, C., *Halo 3: How Microsoft Labs Invented a New Science of Play*. (Featuring R. Pagulayan). *Wired Magazine*, Aug. 2007.
- [15] Ye, J. and Ye, D. HCI and game design: From a practitioner's point of view. *Online Article*, Mar. 2004. <http://www.ye-brothers.com/documents/HCI GAMEDESIGN.pdf>
- [16] Yee, N., Bailenson, J.N., Urbanek, M., Chang, F., Merget, D. The Unbearable Likeness of Being Digital: The Persistence of Nonverbal Social Norms in Online Virtual Environments. In *the Journal of CyberPsychology and Behavior*, 10, 115-121, 2007.