Preliminary Evaluation of the Interactive Drama *Facade*

Rachel Knickmeyer
College of Computing
The Georgia Institute of Technology
rnickm@cc.gatech.edu

Michael Mateas
College of Computing and Literature, Communication & Culture
The Georgia Institute of Technology
michaelm@cc.gatech.edu

ABSTRACT
There is growing interest in technologies that support user experiences emphasizing aesthetic satisfaction and enjoyment rather than task accomplishment. Evaluating such experiences remains an open research problem. Here we describe a methodology for evaluating the interactive drama *Façade*, and present the first experimental results. Interactive dramas are "pure" hedonic experiences, forcing a focus on experience quality rather than efficiency and ease of use. Through the coding of retroactive protocols, we reveal play patterns whereby interaction failures are leveraged into new player goals, thus supporting players in maintaining positive interest in the experience even in the face of interaction failures.

Author Keywords
Retroactive protocol analysis, qualitative evaluation, game evaluation, hedonic evaluation

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Evaluation.

INTRODUCTION
In recent years there has been a growing interest in designing information technologies that support rich and complex user experiences, including satisfaction, joy, aesthetics, and reflection, in addition to task accomplishment. Many traditional HCI evaluation methods, with their frequent focus on efficiency and task accomplishment, are often inappropriate for evaluating the aesthetic and experiential aspects of such systems. While workshops such as Funology [1] have begun to establish theoretical frameworks for the design and evaluation of the hedonic aspects of information technology, evaluation of such experiences is still an open research question. Recent evaluation studies in this area include Hook et. al. [2], which employed interviews and observation techniques to evaluate interactive art, and Stasko et. al. [5], which employed longitudinal studies to evaluate user experiences with ambient, aesthetic information displays. In this paper we describe a methodology for evaluating the interactive drama *Façade* [4], and present the results of the first experiment using this methodology.

Knickmeyer, R., and Mateas, M. 2005. Preliminary Evaluation of the Interactive Drama *Façade*. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI 2005)*. Portland OR, April 2-7. New York: ACM Press, pp. 1549 - 1552, 2005.

In an interactive drama the player enters a virtual world, interacts with computer controlled characters, and through her interaction influences both the characters and the overall development of the story. An interactive drama is in some sense a "pure" hedonic experience, immersing the player in a dramatic social interaction without providing, as most games do, a clear player goal; the player invents goals for herself as the interaction with the characters unfolds. In *Façade* the player visits the married couple Grace and Trip at their apartment, quickly becoming entangled in the high-conflict dissolution of their marriage (reminiscent of an interactive *Who's Afraid of Virginia Wolff*). The player interacts from a first person perspective, moving about the world, manipulating objects, and, most significantly, talking to the characters through unrestricted, typed natural language. Since the player's interaction effects the long-term development of the story, the experience has replay value, in that different interaction approaches will result in different story trajectories. Given the technical and design difficulties of creating real-time, animated, AI-controlled characters that respond broadly and robustly to natural language input, there will inevitably be interaction failures in which the characters respond inappropriately to player interaction. *Façade* was designed to help the player maintain immersion in the experience even in the face of interaction failures. Thus, one of the areas we focus on in our evaluation is the player's response to interaction failure.

Our results reveal a positive overall evaluation from players, that player's are motivated to play again in order to try different strategies and, most interestingly, play patterns whereby even interaction failures are leveraged by the player into new opportunistic goals, thus supporting players in maintaining positive interest in the experience even in the face of interaction failures.

METHODOLOGY
For *Façade*, we define a successful experience as one in which players experience a sense of agency, maintain engagement, and are motivated to replay in order to try different interaction strategies. By agency we mean that players experience both the immediate and long-term

effects of their actions as related to their goals. As interaction breakdowns are likely to negatively impact both immersion and engagement, the player's response to interaction breakdown is a major focus of this study.

We use Retroactive Protocol Analysis as our primary evaluation tool, supplemented with a post-experience interview to gather general player reactions. We chose the retroactive over a standard talk-aloud protocol in order to

Code	Desc	Exp	Code	Desc	Exp
Agency:			Misc:		
LOCAGE	Local Agency	"He's reacted to me picking up the drink"	LAUGH	Player Laughter	
GLOBAGE	Global Agency	"I felt responsible for the outcome"	IRRIT	Unspecified irritation or confusion	
LOSSAGE	Loss of Control / Agency	"I don't know what to do, they aren't listening"	Storyline:		
CHARRESPAP	Appropriate Response	"I liked her response"	BACKGRD	Interest in Background or Continuing story expressed	"I want to know more about their proposal"
CHARRESPNEG	Inappropriate Response	"That response didn't make any sense"	CLOSURE	Sense of closure / end satisfaction expressed	"I liked the ending"
Curiosity:			DISINT	Disinterest in the story expressed	"I'm not really listening anymore to the story"
PHYSEXPL	Exploring Physical Space	"I'm seeing what else I can touch here"	DISSAT	Dissatisfication with outcome	"The ending was very abrupt"
AIEXPL	Exploring AI Structure	"We'll see if the parser understands that"	IRRITATE	General Irritation with storyline / Confusion with storyline	"I don't really understand what's happening anymore"
CHAREXPL	Exploring Character	"I want to hear more about Trip"	Program Bugs:		
Fantasy:			STUCK	Player unable to move	
IMMERSE	Noted feeling of immersion	"This is very socially awkward"	Replay:		
CHARPERS	Acting naturally, or acting as yourself	"I'm doing what I'd normally do in this situation"	DESREPLY	Expressed Desire to Replay	"I want to play again"
FLIRT	Flirtation	"I'm going to hit on Trip"	STRATALT	Specific Strategy Alteration based on previous game	"I'm going to try favoring Grace this time"
INSTIGATE	Instigation / Insulting	"I'm going to try and get them to argue with me"	NOTEDDIFF	Noted Differences in the storyline / actions	"That was neat, it was different from the last game"
SYMP	Sympathy with Characters	"I understand what she's going through"	Representation		
FRUST	Frustration with Characters	"Trip is irritating me now"	REPPOS	Positive reaction to the representation, specifically facial expressions	"I really love the facial expressions, very expressive"
Challenge:			REPNEG	Negative reaction to the representation	"The characters are a little cartoony for me"
STRAT	Indication of a use of specific strategy to created desired outcome	"I'm going to try favoring Trip to see what happens"			

Table 1: Table shows the complete coding scheme used to analyze the video data collected from participants. Codes fall into six categories: Agency, Fantasy, Challenge, Storyline, Miscellaneous, and Representation

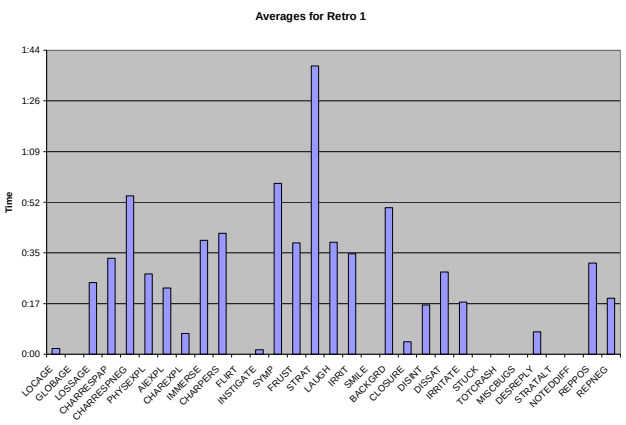

Figure 1. Protocol code avg. across players for first play.

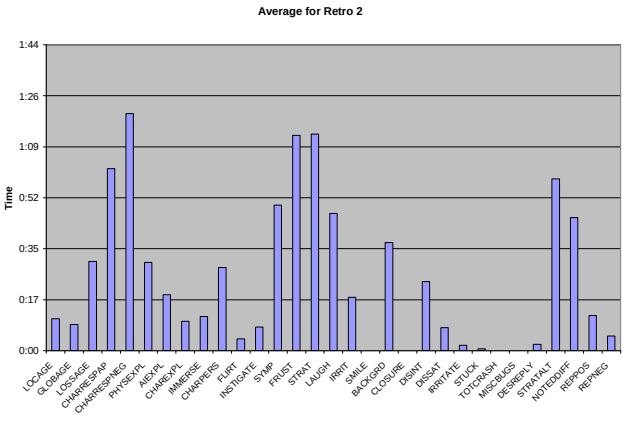

Figure 2. Protocol code avg. across players for replay.

prevent interrupting the fast-paced real-time flow of the experience. In our retroactive protocols, players play Façade while the screen is videotaped. The players then describe their interactions and reactions to the characters in the game while watching the tape and consulting a generated script of their game play. In order to determine *Façade's* replay value, as well as to discover how player's interaction strategies change during replay, they then play *Façade* a second time. The entire protocol is recorded and used for analysis.

We used Malone [3] as a starting point for developing our coding scheme for the retroactive protocols. Malone defines three categories that determine successful game experiences: challenge, fantasy, and curiosity. Our coding scheme expands on this, looking specifically at categories such as Agency, Exploration, Strategy, and Disinterest. The complete coding scheme can be seen in Table 1, including descriptions and example player comments that satisfy each code. Using this scheme, we coded the videos of the retroactive protocol, tagging everything the player said with one or more codes.

RESULTS

Eight players participated in the study, five males and 3 females. There were no gender differences in play patterns or in player's overall evaluation of the experience in the post-interview. Ages ranged from 24 to 38 and gaming experience varied from minimal to quite experienced. Surprisingly, prior gamming experience was not a factor in determining play patterns or overall evaluation of the experience.

GENERAL RESULTS

Six out of the eight players reported during the post-interview that they enjoyed the experience and would like to play the game again. This indicates that, in general, players consider *Façade* to be a successful experience.

We noted two distinct patterns of play. Some players spent the majority of their time exploring and tuning their gaming strategy (STRAT). Other players spent significantly less time on strategy but appeared to be more "immersed" in the experience. They had high scores in IMMERSE and CHARPERS. The first type of player is explicitly trying to manipulate the characters to cause certain outcomes, where the second type of player is trying to "be herself" and "act naturally" within the experience.

Figure 1 shows the average totals for codes in the first retroactive protocol, figure 2 in the replay protocol. The height of the bars represents the length of time players talked about the code category during the retrospective protocol. For example, in figure 1 players spent on average about 25 seconds making comments about the appropriateness of character responses (fourth bar over from left). The total length of time of one play-through of *Façade* varies, as players may take actions that cause the experience to end prematurely (e.g. the player may be thrown out of the apartment for being socially inappropriate too often), or may bring up more or fewer conversational topics. On the long side, a singe playthrough can take 20 minutes.

On replay, we noted a general tendency for players to switch from immersive gameplay (acting naturally) to one involving more strategy. Though the STRAT bar is smaller in figure 2 than in figure 1, the total amount of time the player comments on strategy-based play in the protocol is STRAT + STRATALT; this sum is larger in figure 2 than figure 1. We also found that replay games tend to be longer than the first game, and that the ratio of appropriate to inappropriate character responses (CHARRESPAP to CHARRESPNEG) goes up in replay, indicating that players learn how to evoke more satisfying reactions from the characters. In addition, during replay we see significant activity in STRATALT and NOTEDDIFF, indicating that players actively adjust their interaction strategies in response to the previous experience, and actively note and enjoy the conversational and story-level differences evident during replay.

Patterns in Interaction breakdowns

In order to discern how failures in the interaction affected players' experiences with the system, we graphed each game experience using a Gantt Diagram. This allows us to

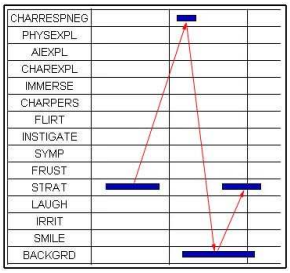

Figure 3

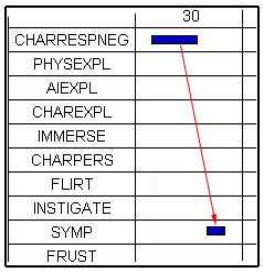

Figure 4

look for temporal patterns in gameplay and compare play patterns across multiple players. This analysis revealed three distinct patterns relating to interaction breakdowns.

Background Interest

Four out of the eight players exhibited this pattern at least once: an interest in the story and character background triggered by an inappropriate character response which then, in turn, triggered a change in player strategy. This pattern is depicted in figure 3. In this case the characters misunderstood (or failed to understand) the player, and yet, in their response, revealed new and interesting back-story information that the player used to opportunistically formulate new interaction goals. This pattern indicates that even when the AI appears to break, the system encourages maintained interest in the storyline and new strategies based on that interest.

Player Affective Response

Every player exhibited the following pattern at least once: an inappropriate character response elicited an affective response from the player to either the characters' personality or background, coded as SYMP or FRUST. This pattern is depicted, using a SYMP example, in figure 4. Note that the affective response FRUST (frustration) does not mean that the player described frustration at the interaction breakdown, but rather described frustration with the character personality or background. We regard this later type of frustration positively, as it is an indicator of character believability and player engagement. This pattern indicates that breakdowns in the system, rather than negatively impacting character believability, sometimes trigger an emotional response to the characters that may increase engagement and believability.

Meta-Play

Additionally, six out of the eight players exhibited a pattern in which inappropriate character responses were followed by a shift in strategy. This pattern is depicted in figure 5. Most of the time this indicated that the players recognized the failure of the AI system and shifted their strategy in order to try and gain the response or outcome they were originally hoping for. In some cases this pattern was merely an indication that the player gave up on one strategy after recognizing the failure and moved onto a different goal entirely, which indicates not true meta-play, but more of an implicit continued interest in the background story. In the first case, this indicates that players are willing to adjust to system limitations in order to achieve their goals and in the second, like the *background interest* pattern, that players opportunistically make use of breakdowns to pursue new interaction opportunities. In either case, the pattern itself indicates that players are willing to adjust to system limitations in order to push the experience further.

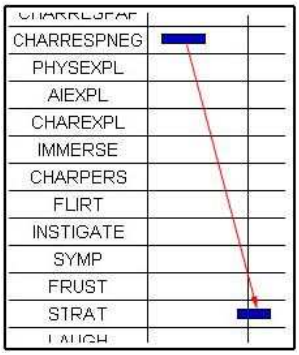

Figure 5

CONCLUSIONS

Façade succeeds as an experience by maintaining player engagement in the face of interaction failures. More often than not, interaction failures actually lead to affective responses or an interest in the background stories of the characters as opposed to a loss of engagement and immersion. *Façade's* design strategy of using autonomous characters that actively move the situation forward, even when the player is misunderstood, may be useful in other autonomous character contexts, such as character-based training environments.

REFERENCES

1. Blythe, Mark, Kees Overbeeke, Andrew Monk, and Peter Wright (Eds.). *Funology: From Usability to Enjoyment*. Kluwer Academic Publishers, 2003.
2. Hook, Kristina, Phoebe Sengers, and Gerd Andersson. Sense and Sensibility: Evaluation and Interactive Art. In *Proc. CHI 2003*, ACM Press (2003), 241 – 248.
3. Malone, Thomas W. Heuristics for Designing Enjoyable User Interfaces: Lessons from Computer Games. In *Human Factors in Computing Systems*, Ablex Publishing (1982), 63 – 68.
4. Mateas, Michael and Andrew Stern. Integrating plot, character and natural language processing in the interactive drama *Façade*, *1st International Conference on Technologies for Interactive Digital Storytelling and Entertainment (TIDSE '03)*, Darmstadt, Germany , March 24 – 26, 2003 .
5. Stasko, John, Todd Miller, Zachary Pousman, Christopher Plaue, and Osman Ullah. Personalized Peripheral Information Awareness through Information Art. In *Proc. UbiComp 2004*, ACM Press (2004), 18 – 35.