User Perceptions of Adaptivity in an Interactive Narrative

Karen Tanenbaum, Marek Hatala, and Joshua Tanenbaum

School of Interactive Arts + Technology,
Simon Fraser University
250-13450 102 Avenue
Surrey, BC, V3T0A3 Canada
{ktanenba,mhatala,joshuat}@sfu.ca

Abstract. We present results from a user study of the Reading Glove version 2.0, a combination wearable and tabletop interactive narrative system. The system was designed to study user perceptions of adaptivity. The system's reasoning engine guides users through the story using three different recommendation modes: random recommendations, story content-based recommendations, and user model based recommendations. We look at the differences in user behaviour and experience across the three recommendation systems, using information from system logs and user surveys and interviews.

Keywords: Adaptivity, User Modeling, Expert Systems, Interactive Narrative.

1 Introduction

This paper presents the results of a study using the Reading Glove, an adaptive narrative system with a combination wearable and tabletop interface. The goal of this research is to turn a critical eye on the notion of adaptivity, specifically within the realm of tangible and ubiquitous systems. In educational and workplace applications, adaptivity is typically task-oriented and aimed at helping users achieve a particular learning or productivity related goal. This means that the adaptive mechanisms can be much more explicit, intervening directly with the user to offer them assistance or advice. In ubiquitous environments, however, the nature of the interaction with technology shifts. Computational elements are embedded in the environment or in smaller, handheld devices. Users may not be paying explicit attention to the system, and the activities taking place are less task-oriented. Some of the most common uses of adaptivity in ubiquitous spaces are for leisure activities, such as museum guide systems that combine entertainment with education, or domestic systems that automate or anticipate common user behaviours. Since users of these systems are less focused on interacting with the technology itself, the goal of the system is to unobtrusively monitor the users and adapt itself to suit them in some way. The novelty of this kind of interaction is a significant issue in constructing adaptive components that work as intended [1]. In adaptive systems outside of the home, such as in museums and other educational settings, this novelty extends not just to the method of interaction but also the frequency of use. Most of the time, these systems

Joseph A. Konstan et al. (Eds.): UMAP 2011, LNCS 6787, pp. 389–400, 2011.

will be "single use", with each person interacting with the system being a new user who needs to learn the interaction paradigm quickly. Unlike domestic or personal systems that have time to learn about their users gradually, ubiquitous public systems have a much more limited window in which to deploy effective adaptation.

The study presented here was designed not to test a particular hypothesis or evaluate the performance of a specific algorithm, but rather to explore the nature of the user experience of adaptivity in this kind of one-time or infrequent use system. We were not concerned whether the adaptive system was "useful" or "effective" in an objective sense, but rather in seeing what kind of sense the users made of the system and of the different forms of adaptive behaviour displayed by the system. We believe this kind of exploratory and subjective experience-oriented research is necessary for truly understanding how and why to design adaptive systems of this nature.

2 Previous Work

2.1 Perception of Adaptivity and Novelty

There are a handful of studies which have taken a similar approach to investigating the user experience of novel systems. Williams et al. focus specifically on the nature of space in intelligent and augmented environments, looking at how people understand ubiquitous computing as a "spatially situated phenomenon" [2]. They created an installation called SignalPlay, which involved a series of large, moveable props which each had a different effect on the soundscape of the room. Visitors to the space had to experiment with the objects in order to understand how they worked; no explicit instruction was given. From observing visitor interactions with the props, the authors identified three modes of object interaction used when learning how to control and interpret new interaction mechanisms: 1) iconic, where they interact in ways suggested by what the props represented; 2) intrinsic, where they interact based on physical characteristics of the objects; 3) instrumental, where they interact based on the effect it has on the system [2].

On a simpler scale, Svanaes and Verplank approach this issue from a different direction with a study of the naturally arising metaphors and mental models that people create when playing with a set of interactive tiles [3]. The authors observed that participants spontaneously made use of five fundamental metaphors: Cartesian space, state space, linear time, relational metaphors (human relations), and paranormal phenomena. Each of these reflects a different mental model that the users were applying to learn how to interact with the system. Understanding these mental models can suggest different ways of designing a tangible interface to leverage the intuitive use of that model, or perhaps to explicitly design against it if the model leads to inappropriate intuitions about how the system works. Both of these studies show how a rich, qualitative-focused investigation of how people interact with computational systems can result in insight into the nature of these technology-enhanced experiences and how to design them.

2.2 Interactive Narrative and Recommender Systems

Most work on intelligent narrative systems centers around how to adapt the story and environment to choices made by the interactor, i.e. how to restructure the plot so that story coherence is maintained or how to create non-player characters that can interact with the player in a life-like manner [4]. In the Reading Glove system, the narrative and the environment are fixed. The interactor selects what order the story is heard in, but cannot fundamentally change what happens. The reasoning engine that drives the guidance system on the tabletop thus functions essentially as a knowledge-based recommender, helping the "reader" move through the story in a coherent manner. As a result, the intelligence techniques used in the system are most similar to those used in recommender systems in educational and informational applications, where the goal is to present a static body of content to the user in an intelligent and dynamic manner based on her choices and actions [5-7].

3 The Reading Glove System

The current Reading Glove system is version 2.0 of an earlier iteration of the project. The first version, discussed in [8-9], consisted of a glove-based reader and a set of tagged objects used to access a non-linear story. The current version adds an intelligent recommender system and tabletop display (See figure 1). These additions assist interactors in navigating the narrative while also allowing us to study user perceptions of adaptivity. This paper is the first to discuss results from this iteration.

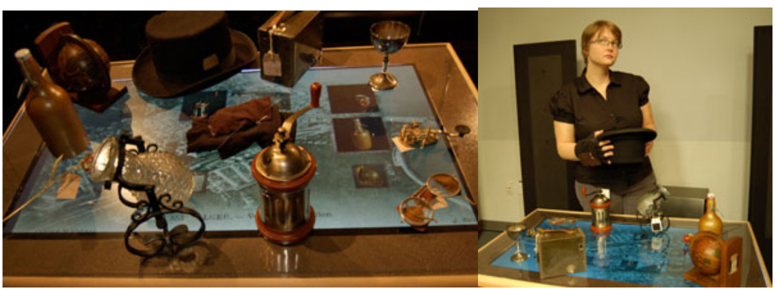

Fig. 1. The objects on the tabletop (left) and a reader using the system (right)

3.1 The Reading Experience

Interaction with the Reading Glove system is straightforward. The "reader" puts on a soft fabric glove and begins by picking up one of the objects sitting on a tabletop. This tabletop displays pictures of each object arranged in a rectangle (see figure 3 below). When the palm of the glove registers the tag on the object, a segment of recorded audio narration is played back over the speakers. Several seconds before the clip ends, the tabletop display delivers a set of recommendations on which object to pick up next by enlarging and brightening photos of the recommended objects. The reader can choose to follow the on-screen advice or not. Each object has two clips of audio narration associated with it, so the reader must engage with each object multiple times to uncover all the story fragments.

The story embedded in the Reading Glove system was developed based on the objects, which were picked to fit a certain historical aesthetic. Other aspects of this aesthetic are echoed in the background image of the tabletop display and in the table itself. The plot of the story revolves around a British spy operating in French-occupied Algiers around the turn of the 20th century. The narrative traces the spy's discovery that his cover has been blown and his unraveling of how this came about. The uncovering of facts in the narrative mimics the uncovering of story fragments that the readers perform with the objects. Thus the puzzle-like nature of the story and the interaction support and reinforce each other. The story can also be experienced in a small group, with one person wearing the glove and the others assisting in untangling the narrative and selecting the next objects to engage with.

3.2 The Glove and Objects

The central component of the system is the Reading Glove itself, a soft fabric glove containing an Arduino Lilypad microcontroller, an Innovations ID-12 RFID reader, and an Xbee Series 2 wireless radio. Interactors pick up objects associated with the story, each of which has been tagged with an RFID chip. When the RFID reader in the palm of the glove detects a tag, the tag ID is communicated wirelessly via the Xbee radio to a second Xbee unit connected to the serial port of a laptop. The serial data is read into a java program in Eclipse which processes the tag activation and triggers the audio playback of a specific "lexia": a pre-recorded story fragment associated with the object.

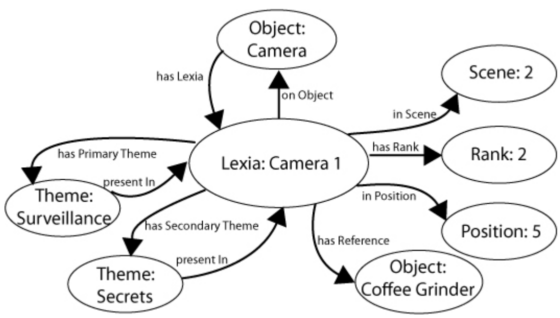

Fig. 2. The structure of the ontology for one lexia

3.3 The Table and Reasoning Engine

In addition to generating audio feedback, picking up an object also triggers the reasoning engine to generate a set of recommendations that will be shown to the interactor when the audio clip nears its completion. The reasoning engine is a rule-based expert system written in the Jess language. The reasoning component relies on an OWL (Web Ontology Language) ontology that encodes semantic knowledge about the story content. A rule-based expert system was used because this type of artificial intelligence mostly closely mimics the behaviour of a human expert, in our case the interactive story writer, as suggested by [10]. The rules can be hand-crafted based on expert knowledge and thus do not rely on a large corpus of data from which to generate models or rankings.

Ontology. The ontology has 5 classes and 11 object properties that link classes together in a directional relationship. The object and lexia classes have a reciprocal relationship, with each item in the object class (e.g. the physical object Telegraph Key) linking to two entities in the lexia class (e.g. the sound files Telegraph Key 1 and Telegraph Key 2) and each lexia connecting back to the object. See figure 2 for an example of a specific lexia in the ontology, Camera 1, on the object Camera. The lexia class also has a set of non-reciprocal object properties connecting each sound file to different pieces of information. The "hasRank" property indicates how important the lexia is to the overall narrative, as determined by us as the story authors. Rank varies from 1 to 9, with 1 being the most important. The "inScene" property indicates what scene each lexia was part of; there were 4 scenes determined by changes in the location of the narrative. The "hasReference" property was only active for some lexia, those which contained a direct reference to another object within the text of the audio clip. For example, the camera1 lexia includes the sentence "I made certain to lose myself in the chaotic traffic of one of the city's open air markets before stopping to inspect the coffee grinder.", so in the ontology the lexia is linked to the coffee grinder object. Finally, each lexia is associated with 2-3 themes present in the story, such as "surveillance" or "disguise". This relationship was also represented reciprocally between the lexia and theme classes with the properties "hasPrimaryTheme" and "hasSecondaryTheme" connecting lexia to themes and "presentIn" connecting theme to lexia. All of the relationships in the ontology were asserted as facts in the JESS rule base at the start of each system run. For the implementation details on working with ontologies in Jess see [11].

Fig. 3. The tabletop screen in neutral (left) and recommender (right) states

Recommendations. The Jess rules use this knowledge base to recommend a set of three objects that will be most likely to advance the interactor's understanding of the story. Thus the recommender system acts as a kind of "expert storyteller", leading the reader through the narrative. The recommendations appear on the table several seconds before the end of the lexia. This delay is intended to focus attention on the story and objects rather than the display, encouraging the user to listen to the full lexia rather than just skip ahead. During most of the lexia playback, all 10 objects are visible on the screen in small, semi-transparent boxes. When the recommendation system kicks in, the pictures of the recommended objects grow in size and become fully opaque (See figure 3). The display remains in this state until another object is picked up, at which point in reverts to the neutral state.

3.4 Recommender Types

Three separate versions of the recommender were developed: a story content recommender, a user model recommender, and a random recommender.

Story Content Recommender. The story content recommender uses encoded knowledge about the narrative to recommend three objects that will be most likely to continue the story in a coherent and helpful way. The interactor can choose any object to start the story, after which the recommendation system begins to assist based on their ongoing choices. Each of the three recommended objects are chosen based on a different set of criteria: Theme, Importance, or Position. The last lexia chosen by the interactor is used as a “seed” to the recommendation system, generating a set of weights that rank all other available “candidate” lexia. The highest ranked candidate after all the weights are calculated is the one recommended for each criterion.

Theme. The Theme criterion uses the ontology-encoded themes of the seed to evaluate the candidates based on how closely their themes matched. Each lexia has two themes, primary and secondary. The weighting of the candidates is based on whether both the theme and the theme type match the seed. Table 1 gives the weights for ranking seed and candidate themes. If the seed lexia text contains a direct reference to the object of the candidate lexia, this contributes an additional 50 points. After all the weights are calculated and summed together, the candidate with the highest sum is designated the Theme recommendation. Again, for implementation details of applying our weighting scheme using rules we refer the reader to [11].

Table 1. Weightings for matching themes

Seed \ Candidate	Primary Theme	Secondary Theme
Primary Theme	50	20
Secondary Theme	30	40

Position. The Position criterion looks at the chronological order of the lexia and favors candidates that would either move the story forward or fill in the backstory. The highest weights are given to candidates that are 1-4 positions past the seed, while medium weights are given to candidates positioned prior to the seed location, and low weights are given to candidates 5 or more ahead of the seed. So if the seed lexia is in position 5, the candidates in positions 6 would have a weighting of 50, 7-9 would be weighted 30, 1-4 would be weighted 20 and 10-20 would be unweighted. This prioritizes continuity of the story and deprioritizes leaping ahead to the end of the narrative. The candidate with the highest weight at the end of this calculation would be designated the Position recommendation.

Importance. The importance criterion looks at what the most important pieces of the story are and favors recommending the most crucial information. The importance weights combine information about what scene the fragment is in and what the overall rank of each lexia within the scene is. Candidate lexia in the same scene as the seed lexia are given a weight of 50 while candidates from different scenes are unweighted. Next, importance weightings are assigned based on rank, with rank 1 = 45, 2 = 40, 3 = 35, and so on down to rank 9 = 5. The ranks of both of the lexia on an object were summed together with the scene weighting for each candidate lexia. This mechanism was necessary in order to uncover lexia on objects that had not yet been interacted with. For example, an object might have a lexia with rank 8 as the initial state and a lexia with rank 2 as the secondary state. Although the second lexia is very important, if the first lexia is never listened to, the other one will never become available. Summing the importance for both lexia on the object allowed unimportant lexia to be recommended in order to get access to the more important pieces also on the same object. The scene and rank weights were summed and the candidate with the highest sum would be designated the Importance Recommendation.

After all these calculations are completed, the recommendations generated by each of the criteria are presented to the user on the tabletop. Each recommendation has a subtly colored border indicating which criterion it represents, with blue for theme, green for position, and red for importance.

User Model. The user model recommender is built on top of the story content recommender, adding additional weights based on the specific actions the user takes with the system. It promotes lexia that have not yet been listened to by adding weights to the candidate calculations described above. The user model also tracks which of the recommendation streams are followed if the user selects from one of the three highlighted objects. If the user consistently follows one recommendation criterion over the others, the user model component will begin to push that recommendation to the user earlier, before the other two.

Random. The random recommender is simple and straightforward: three objects are selected at random from the set of available objects using a random number generator in Processing, and are presented to the user via the tabletop display. The colored borders around the pictures are maintained, but are essentially meaningless.

4 User Studies

We designed our user study to investigate the following questions using the Reading Glove system:

1. How do interactors respond to the adaptive system?
2. How do the responses differ across the different types of adaptivity?

Our goal was to explore the user response to adaptivity rather than to evaluate the strict effectiveness of the adaptive mechanisms. We were interested in how the users made sense of a system that responded to them in intelligent or intelligent-seeming ways when they were not given any explicit information about what the system would be reacting to.

4.1 Study Protocol

In the fall of 2010 we conducted a mixed-methods user study with 30 participants in roughly one hour long sessions. We collected a wide variety of data, including pre- and post-interaction surveys, a post-interaction interview, video of the participants using the system, and log data generated by the system itself. Participants were randomly assigned to one of three conditions, corresponding to the three versions of the system described above. They were given a brief tutorial on how to use the glove by interacting with a set of training objects, and then engaged with the full system. They were not told which condition they were in, and the only description they were given of what the system did was as follows: “You will be interacting with this collection of objects. Interact with them until you feel like you understand the story. The images on the screen can help guide you through the story. You are free to handle, play with, and move the objects around as much as you like. You may take as long as you like. Let us know when you are ready to stop.”

4.2 Participant Demographics

Of the 30 participants we ran in the study, we had 19 men and 11 women. Ages ranged between 23 to 55 years old, with the median at 31 years. All were graduate level students, 20 working on their Masters degrees and 10 working on PhDs. Most were from media and technology oriented programs. Participants were asked to self-rank themselves on their English fluency, with 18 reporting to be native speakers, 7 reporting as fluent speakers and 5 as advanced speakers. All participants were administered a listening comprehension test at the start of the session as well, to check for English comprehension issues, and all passed.

5 Results and Discussion

To begin to answer our research questions, we looked at descriptive and correlational statistics drawn from the system logs, questionnaires, and interviews. Portions of the interview data were coded in order to generate ratings for how well participants understood the story.

5.1 Basic Response to the System

At the very end of the user study session, we asked participants to fill out a short Likert-style survey consisting of 8 questions. The questions were paired as negative and positive versions of 4 basic concepts, with participants asked to rate them on a 5-point scale consisting of “Strongly Disagree”, “Disagree”, “Undecided”, “Agree”, “Strongly Agree” and “No Answer”. The pairs were presented in a jumbled order to the participant, but are listed here by concept for ease of reading.

- Ease of Use: The system was easy to use; The system was hard to use
- Enjoyment: The system was enjoyable to use; The system was confusing to use
- Story: The actions I took didn't influence the story; My actions changed the story
- Experience Again: I would not be interested in experiencing another story like this; I would like to experience another story this way

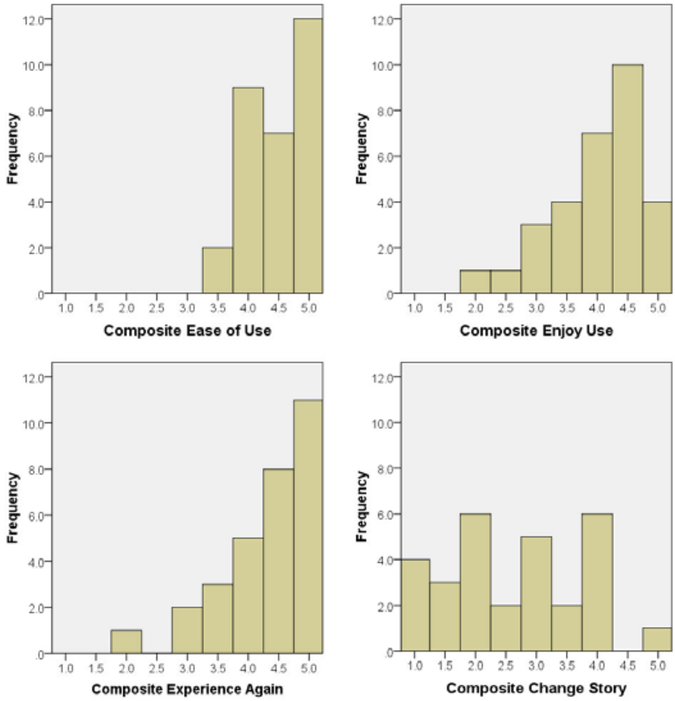

Fig. 4. Frequency distributions for the post-questionnaire items

The charts in figure 4 show the composite participant responses on the post-questionnaire, with “Strongly Disagree” coded as 1 and “Strongly Agree” as 5. Responses from the paired questions were combined to yield a composite measure. The second version of the question was subtracted from 6 (i.e. a score of “5” became a score of “1”) and then the responses were summed and divided by 2. The scores on “Ease of Use” (mean 4.483), “Enjoyment of Use” (mean 4.017) and “Experience Again” (mean 4.317) were consistently high enough that we feel safe in concluding that there were no serious usability issues that were affecting the way participants engaged with the system. The most variable results come from the question regarding whether or not they perceived the story as changing as a result of their actions (mean 2.638). A full discussion of these results addresses separate research questions regarding the interpretation of interactive narrative experiences, and is outside the scope of this paper. No significant correlations were found between the condition that the participant was in versus the ratings they gave the system in the post-questionnaire.

In the post-interaction interview, we asked participants to articulate their understanding of how the system worked. There were three main questions which elicited responses about the adaptive component:

1. Can you describe in your own words how the system worked?
2. What did you think was going on with the pictures of the objects on the screen?
3. How did you decide which object to pick up next?

Nearly all of the participants explicitly stated in the interview that they believed the system was responding intelligently to their actions by highlighting particular elements on the screen, but none of them had any confident guesses about how the system was making the decisions. The most common guess was that the system recommended the chronologically next item; some participants also speculated that it recommended objects that hadn't been picked up yet. One participant described their experience as follows: "So for the first half I picked objects up in sequence according to what was highlighted…The one time I did it out of sequence I got something out of sequence and so it was a little confusing. So there was sort of good bread crumbs there that when I didn't follow them, you know, broke the story a little bit." Similarly, another participant said the recommendation system "was helpful I think in sort of helping me move through the story: when I sort of went away from that I found the story got a little more broken apart and all over the place." We were surprised to find that almost no one recognized that the three recommendations could be distinguished from each other by their colored border; it's possible that the color tints were too subtle. Most people selected from amongst the three recommended objects based on whim or personal attraction to a particular object. One of the participants in the random condition volunteered that sometimes the recommendations "didn't really make sense", but when asked later what he thought the system was doing with the enlarged pictures on the screen claimed that there "I think there's some sort of really complicated algorithm in the background that's figuring out what to display." These statements suggest that participants are strongly inclined to believe that the system is smart—possibly even smarter than them—in that they assume they would not be able to guess at the complexity underlying the system behaviour.

5.2 Response to Different Forms of Adaptivity

After investigating the overall response to the adaptive system, we focused on the research questions stated above and studied the difference between the versions. We wanted to see if there were patterns in participant behaviour that indicated an unconscious reaction to the nature of the intelligence underlying the system, even if they could not articulate that understanding when questioned. We began by examining descriptive statistics based on the data in the system logs, which included elements like how many distinct lexia each person listened to, how many times they followed a recommendation, and how much overall time they spent interacting with the system.

One thing that we noticed early on was that one of our 30 cases was an outlier in almost every metric collected. Because this participant's numbers were so far outside the cluster of everyone else, we chose to discard that data rather than allow it to skew the results. It seemed clear that this person interacted with the system in a very different way than the rest of the participants. If we had a larger sample, it might be possible to identify whether this represents a particular subset of the population who consistently responds in a particular way, but in the absence of more data we cannot tell what is going on.

While examining the descriptive statistics and graphs of the data collected, we noticed that on two key behavioral factors, the participants in the random condition appeared to be on the low end of the scale compared to the participants in the two intelligent conditions (see figure 5). These were “Average Listens per Lexia” and “Total Lexia Activated”, measures that are related to each other. Both of them give an indication of how much of the story was listened to. Since there were 20 lexia, participants who listened to fewer than 20 total lexia obviously did not hear everything. Average listens gives a similar indication of the saturation of the reading, with a score of 1 indicating that they listened to each lexia once, higher numbers showing that they listened to some of the lexia repeatedly, and lower numbers indicating that they did not hear every piece of the story.

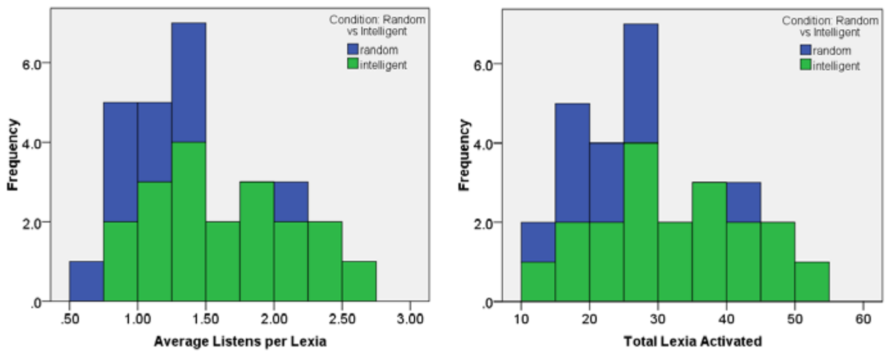

Fig. 5. Frequency distributions for two behavioural measures separated by condition (random = blue, intelligent = green)

We ran an ANOVA on these two factors to see if the apparent correlation between condition and behaviour was significant. There was a significant effect of condition on the number of lexia interacted with: $F(1, 27) = 4.736$, $p<.05$, $w = .33$ as well as a significant effect of condition on average number of listens per lexia, $F(1, 27) = 5.838$, $p<.05$, $w = .38$. What is particularly interesting about this result is that we also ran an ANOVA on amount of time spent with the system, and failed to find a significant correlation between time spent and condition. So it was not simply that the interactors in the intelligent condition spent more time with the system, but rather that they listened to more lexia repeatedly within the time that they spent. This points to a deeper and more dedicated engagement with the system that is driven by the adaptivity.

6 Conclusions and Future Work

We have described a tangible interactive narrative system that uses intelligent techniques to recommend paths through the story space. We presented a user study where participants were exposed to one of three different recommender systems to explore the user experience of adaptive systems. We have shown that users are unlikely to be able to articulate their understanding of the intelligence of the system, but that this doesn’t mean they don’t respond to it at some level. Statistical results regarding behaviour with the system suggest a deeper level of engagement as a result of the adaptive behaviour.

Acknowledgments. We gratefully acknowledge that this project was funded by the PLAYPR group of the GRAND NCE and NSERC Discovery Grants Program.

References

1. Edwards, W.K., Grinter, R.E.: At Home with Ubiquitous Computing: Seven Challenges. In: Abowd, G.D., Brumitt, B., Shafer, S.A.N. (eds.) UbiComp 2001. LNCS, vol. 2201, pp. 256–272. Springer, Heidelberg (2001)
2. Williams, A., Kabisch, E., Dourish, P.: From Interaction to Participation: Configuring Space through Embodied Interaction. In: Beigl, M., Intille, S.S., Rekimoto, J., Tokuda, H. (eds.) UbiComp 2005. LNCS, vol. 3660, pp. 287–304. Springer, Heidelberg (2005)
3. Svanaes, D., Verplank, W.: In Search of Metaphors for Tangible User Interfaces. Designing Augmented Reality Environments, Elsinore, Denmark (2000)
4. Thue, D., Bulitko, V., Spetch, M., Wasylishen, E.: Interactive Storytelling: A Player Modelling Approach. In: Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE), Stanford, CA, pp. 43–48 (2007)
5. Damiano, R., Gena, C., Lombardo, V., Nunnari, F., Pizzo, A.: A stroll with Carletto: adaptation in drama-based tours with virtual characters. User Modeling and User Adaptive Interaction 18, 417–453 (2008)
6. Hatala, M., Wakkary, R.: Ontology-Based User Modeling in an Augmented Audio Reality System for Museums. User Modeling and User Adaptive Interaction 15, 339–380 (2005)
7. Hatala, M., Tanenbaum, K., Wakkary, R., Muise, K., Mohabbati, B., Corness, G., Budd, J., Loughin, T.: Experience Structuring Factors Affecting Learning in Family Visits to Museums. In: Cress, U., Dimitrova, V., Specht, M. (eds.) EC-TEL 2009. LNCS, vol. 5794, pp. 37–51. Springer, Heidelberg (2009)
8. Tanenbaum, K., Tanenbaum, J., Antle, A., Seif El-Nasr, M., Hatala, M.: Experiencing the Reading Glove. In: Tangible, Embodied and Embedded Interaction. ACM Press, Madeira (2011)
9. Tanenbaum, J., Tanenbaum, K., Antle, A.: The Reading Glove: Designing Interactions for Object-Based Tangible Storytelling. In: Augmented Human, pp. 132–140. ACM Press, Megeve (2010)
10. Darlington, K.: The Essence of Expert Systems. Prenctice Hall, Harlow (2000)
11. Hatala, M., Wakkary, R., Kalantari, L.: Ontologies and Rules in Support of Real-time Ubiquitous Application. Journal of Web Semantics 3(1), 5–22 (2005)