Engagement in Interactive Digital Storytelling: Sampling Without Spoiling

Sergio Estupiñán, Kasper Ingdahl Andkjaer, Nicolas Szilas

HAL Id: hal-02128597
https://inria.hal.science/hal-02128597

Submitted on 14 May 2019

Engagement in Interactive Digital Storytelling: Sampling without spoiling

Sergio Estupiñán, Kasper Ingdahl Andkjaer, Nicolas Szilas

TECFA-FPSE, University of Geneva, Bd. du Pont-d'Arve 40, 1205 Geneva, Switzerland
{sergio.estupinan,kasper.andkjaer,nicolas.szilas}@unige.ch

Abstract. Interactive Digital Storytelling (IDS) enables users to influence the unfolding of a story at the plot level using Artificial Intelligence techniques for generating and balancing narrative paths on the fly. Despite their promise, most IDS systems are still unable to consistently deliver engaging user experiences. To further understand the characteristics of this new media, particularly user engagement, we apply the concept of Continuation Desire to an existing Interactive Narrative to dynamically sample engagement during a play session. We use a comparative study to assess the effects of interrupting users during gameplay. This study found no evidence that, if done properly, interrupting the experience spoils it. We find no significant impact on the desire to play again as a result of introducing interruptions either.

Keywords: Interactive Digital Storytelling, Evaluation, Engagement, Continuation Desire, Disruptiveness, Replayability, User Research.

1 Introduction

Interactive Digital Storytelling (IDS) is a relatively new media whose rationale is to empower users to influence the course of narrative events of a computer-based system [7, 8]. The unfolding narrative is dynamically generated using Artificial Intelligence (AI) for narrative generation and balancing. In IDS, users would 'feel' the unfolding story as being their own, since it is the outcome of their own actions that drives it.

The high degree of freedom, increasing complexity, and range of possible narratives, makes it difficult to keep the structure of the narrative engaging and meaningful. Such variability is also problematic from an evaluation perspective. The focus of evaluation must be on the Interactive Narrative Experience (INE), not the final product.

Current evaluation of IDS systems has mainly focused on the global appreciation of the INE inquiring users about their experience after a play session using surveys and standardized questionnaires. There is, however, a lack of research in the IDS community into the factors that determine, during run time, if and why an IDS system is perceived as meaningful and engaging by the users at a given time.

In game studies, one such concept is engagement [11], which has been used to explain the user involvement with the system, albeit not specifically with a narrative focus. Engagement theories have previously been used to measure games as processes,

and given their narrative focus, they could fit well with IDS. This article introduces a method inspired by engagement research and performs a comparative study to measure the impact of such a method on users.

2 Related Work

Early research on the evaluation of IDS has provided basic results beyond the mere display of output examples [3] and the evaluation of the product. Ad-hoc questionnaires have been used in an attempt to evaluate the general qualities of systems (e.g. eval. B in [1]). Another line of research used objective metrics to measure the user experience (e.g. aggregating interaction logs). Qualitative studies have also been conducted based on interviews and content analyses [12]. In the video games domain, related research has focused on the evaluation of in-game metrics as well as fine-tuning parameters within games [13]. Log analyses [5] and interrupting questionnaires have also been used, although not necessarily in relation to the narrative dimension. It is only with the IRIS project [6] that a IDS-specific evaluation framework was proposed, focusing on the general assessment of users' Interactive Narrative Experience after using an IDS system [9].

While all of the above self-report approaches are *a posteriori*, there are other methodologies which have proceeded differently. To tackle the problem of measuring engagement in a condensed way, Continuation Desire (CD) attempts to perform sampling during an experience: users are interrupted and asked to quantify to which extent they want to continue the experience, what they want to do next and why they want to continue [10, 11]. This approach produces a continuous CD curve and accounts of the narrative, making it possible to trace the evolution of CD as the Interactive Narrative unfolds.

The multiple facets associated with the term engagement, including flow, immersion, and enjoyment, share this common trait: the willingness to continue taking part in the activity, which is what Continuation Desire measures. It has therefore been argued that CD acts as an indicator of engagement [11], although the relation is one-way.

This answer to the problem of *a posteriori* approaches does not come without problems however: Injecting questionnaires into an experience could disrupt and bias the experience, which should be avoided [10]. This study addresses this problem.

3 Engagement in Interactive Digital Storytelling

Given the unsuitability of relying only in non-disruptive or psychophysiological measures for engagement research, the challenge becomes to minimize the disturbance of more intrusive methods so as not to bias the results. The scope of this article is limited to designing and implementing an in-game Continuation Desire evaluation approach and to measure the disruption caused. We believe that such a method should meet these requirements: 1) Gather fine-grained data during the interaction with an IDS system, and 2) The measuring instrument should not bias the results.

4 Experiment Design

To test the hypothesis that introducing interruptions during the experience will not have a significant negative impact on the experience, we implemented a CD questionnaire into the IDS system "Nothing For Dinner"[1] and set up a comparative study using self-reporting metrics disruptiveness and replayability desire. Two independent groups were recruited using a scientific crowdsourcing platform [2]: Interruption group (N=35), and Control group (N=42). There were 36 males and 41 females, 18–40 years old (M=27.3 y/o, SD=5.95 y/o), speaking English as a first language. There was no recruitment requirement for computer proficiency or gaming profile. Compensation was 1.50 £ for around 23 minutes, the study was taken online on their own computers.

4.1 Metrics

The content of the interruption questionnaire is not evaluated in this article, it is mentioned as a proposal for a way to measure user appreciation of the INE. The relevant metrics for this experiment are: 1) *Disruptiveness*; how much users felt the interruptions were in the way of the experience (measured in the Interruption group), and 2) *Replayability desire*; how willing the user was to try the experience again (sampled in both groups).

4.2 Interruption Design

Triggering. The moment of interruption must be chosen to avoid generating user discomfort (e.g. while the user is reading or selecting an action) and also since actions may occur in parallel in the system (e.g. agents may decide to interact with each other or with the player). We implemented the following interruption triggering algorithm:

```
1)Play session starts
2)Wait 'x' seconds
3)Wait until all User-Initiated Actions (UIA) are completed
4)Watch for new UIA
5)Wait until UIA is completed
6)Wait for system dialogs to end
7)Wait 'y' seconds
8)Is there any active UIA?
YES) Go to 6
9)Trigger interruption, pausing the system
11) Has the interruption appeared 'z' times? NO) Go to 2
12) Wait for 't' seconds
13) End
```

In step 5, the algorithm waits for all User-Initiated Actions (UIA) to finish so the user can see the outcome of their actions prior to each interruption.

---

[1] http://nothingfordinner.org/

[2] https://prolific.ac/

User Interface. We consider that the UI design should meet the following characteristics: 1) Users can observe the state of their session while the questions are asked, so they feel they are in the system (non-masking interface), and 2) The elements giving questions, as well as users' means of answering, should mimic the existing system to ensure continuity between system and sampling.

4.3 Questionnaire Contents

Continuation Desire. The Engagement Sampling Questionnaire (ESQ) [11] was presented in each interruption: 1) "Do you want to continue the experience?" (Likert scale), 2) "What makes you want/not want to continue the experience?", and 3) "What do you want to do next in the game?". The last question is relevant from a narrative perspective since it assesses the audience's anticipatory behavior, a core property of narratives [4].

Affective state. Users reported their affective state during each interruption using the Self-Assessment Manikin (SAM) [2], a validated cross-cultural affective assessment tool in three dimensions (Valence, Arousal, Dominance). Chosen for its fast learning curve and rapid reporting, users were shown SAM in the preliminary phase of the study. The affective results are not in the scope of this article.

4.4 Procedure

Upon being informed about the financial compensation and accepting the research consent, participants started the experiment by being redirected from the scientific crowdsourcing website to the study website. The procedure was:

1. Training questions (Interruption group): SAM is introduced and explained.
2. Anticipation: 1) "I want to begin playing" (Likert scale). 2) "What makes you want or not want to begin playing?" (freeform text).
3. The experience is presented in full. When the interruption procedure has run its course or time is reached for the Control group, the post-questionnaire is presented.
4. Post-game question.: "I want to try the game again" (Likert scale).
5. Regarding the interruptions (Interruption group): "The in-game interruptions spoiled my gaming experience" (Likert scale), "Overall, filling out the interruption questions was:" (Very difficult - Very easy, Single Ease Question)

We used seven-point Likert scales since they provide a good balance between degrees of freedom and ease of use. Participants received the compensation at the end.

5 Results

Interruption distribution:
From the start of the experience to *Interruption 1*, two clusters were found: C1, size = 20, 5.63 minutes elapsed in average (SD=51.32 s), results interval [222, 400] s, and C2 2, size = 15, 8.50 minutes elapsed in average (SD=63.9 s), results interval [426, 650] s.
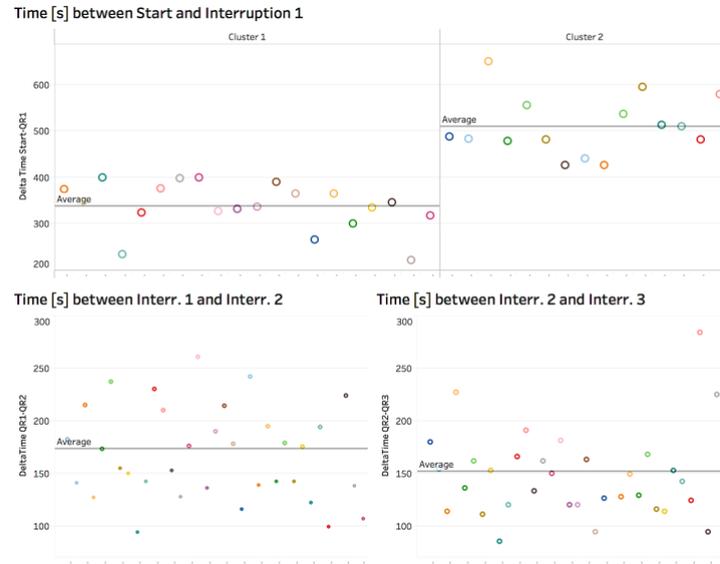
Figure 1. Elapsed time between Question Rounds in seconds.

Next, 97% of *Interruption 2* triggered around 2.90 min in average (SD=55.52 s) interruptions being evenly distributed around 94 and 261 s after the first interruption. Finally, 94% of *Interruption 3* triggered around in average 2.54 min after *Interruption 2* (SD=51.73 s), uniformly spaced around 85 and 227 seconds.

Disruptiveness: Participants in the Interruption group had a neutral opinion of interruptions having spoiled their experience, with an average of 4.11 (SD=1.61), 95% Confidence Interval ranging from 3.57 to 4.63 (4 = Neutral, 5 = Agree a little).

Difficulty: The perceived difficulty filling out the interruption questionnaire was calculated using the Single Ease Question, scoring 5.79 (Moderately easy) in average (SD=1.12). The calculated 95% Confidence Interval ranged from 5.42 to 6.16, where 4 corresponds to Neutral, and 7 to Very easy.

Replayability desire: The Control group scored 3.07 whereas the Interruption group 2.82 (Disagree a little) in average. An independent two-tailed t-test was performed, finding no significant difference (p = 0.536). The calculated Cohen's d is 0.1362, so the effect size is very small.

6   Discussion and Conclusion

The main objective of this article is to assess the disruptiveness of the proposed approach for measuring Continuation Desire (an indicator of the presence of engagement) within the Interactive Digital Storytelling domain. A comparative experiment was made using the IDS system "Nothing For Dinner" with an Interruption group (N=35) and a Control group (N=42), each session lasted around 23 minutes.

The obtained results indicate that the interruption triggering system performed consistently, spreading the interruptions evenly over the entire session. More importantly, it was found no major perception of disruptiveness when using the proposed approach for the sampling of Continuation Desire. In terms of replayability, we found no statistically significant effect when introducing the interruptions. The content of the interruption questionnaire was not analyzed in this article, it remains the basis for future work assessing its usefulness compared to existing evaluation approaches. These findings support the notion that in-experience sampling could be performed without creating major impact on users. It remains our long-term goal to advance the evaluation of IDS, for which we consider this study a first successful step.

References

1. Barber, H., Kudenko, D.: Generation of adaptive dilemma-based interactive narratives. IEEE Trans. Comput. Intell. AI Games. 1, 4, 309–326 (2009).
2. Bradley, M., Lang, P.J.: Measuring Emotion: The Self-Assessment Semantic Differential Manikin and the. J. Behav. Ther. Exp. Psychiatry. 25, I, 49–59 (1994).
3. Cavazza, M., Charles, F., Mead, S.J.: Characters in Search of an author: AI-based Virtual Storytelling. In: Balet, O. et al. (eds.) International Conference on Virtual Storytelling (ICVS 2001). LNCS 2197. pp. 145–154 Springer, Heidelberg (2001).
4. Eco, U.: Lector in fabula: Le rôle du lecteur ou la Coopération interprétative dans les textes narratifs. Grasset (1985).
5. Kim, J.H., Gunn, D. V, Schuh, E., Phillips, B.C., Pagulayan, R.J., Wixon, D.: Tracking real-time user experience (TRUE): a comprehensive instrumentation solution for complex systems. CHI '08 Proc. SIGCHI Conf. Hum. Factors Comput. Syst. 443–451 (2008).
6. Klimmt, C., Roth, C., Vermeulen, I.: The UX-tool: Measurement device and documentation for empirical assessment of the UX in IS. January 2012, (2012).
7. Molnar, A., Kostkova, P.: Interactive Digital Narrative. Routledge (2015).
8. Murray, J.H.: Hamlet on the Holodeck: the future of narrative in cyberspace. MIT Press (1997).
9. Roth, C.: Experiencing Interactive Storytelling. Ph.D. thesis. VU University Amsterdam (2015).
10. Schoenau-Fog, H., Birke, A., Reng, L.: Evaluation of Continuation Desire As an Iterative Game Development Method. In: Proceeding of the 16th International Academic MindTrek Conference. pp. 241–243 ACM, New York, NY, USA (2012).
11. Schoenau-Fog, H.: Hooked! - Evaluating engagement as continuation desire in interactive narratives. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 219–230 (2011).
12. Seif El-Nasr, M., Milam, D., Maygoli, T.: Experiencing interactive narrative: A qualitative analysis of Facade. Entertain. Comput. 4, 1, 39–52 (2013).
13. Sundstedt, V., Bernhard, M., Stavrakis, E., Reinhard, E., Wimmer, M.: Game Analytics. (2013).