RECORD, REPLAY & REFLECT – A FRAMEWORK FOR UNDERSTANDING (SERIOUS) GAME PLAY

Anton Eliëns [1,2] and Zsófia Ruttkay [2]

[1] Intelligent Multimedia Group, VU University Amsterdam
[2] Creative Technology, University of Twente

KEYWORDS
serious games, learning, machinima

ABSTRACT

In this paper we set the first steps towards defining a framework for understanding game play based on recording user actions allowing for later replay, which we may characterize as symbolic machinima. After discussing issues of user tracking in virtual environments, we introduce a behavioral model for game play, discuss its relation to existing game reference models, and define a metrics to characterize behavioral discrepancy, for example from a norm scenario in a management game. Finally, we will indicate the possible benefits of our approach for understanding individual users' game play, as well as for the re-use of analysis data for authoring alternative game scenarios.

INTRODUCTION

Educators continuously face the problem of motivating their students to learn, to consult textbooks, do exercises, reflect on the material, and repeat this cycle over and over again. Although e-learning technology made the learning materials more readily available, as well as additional facilities for searching information, making rote-learning apparently obsolete, it did clearly not help in solving the motivational issues, that is the students' involvement with the subject, the time spent on studying and solving problems, and the unavoidable chore of repetition, that is inherent in all learning.

Recently, serious games have come to the rescue, allowing for a more active form of learning, Gee (2003), in a wide area of domains, including language learning, management, as well as behavioral attitudes, for example with respect to climate change, Eliens et al. (2007b). Yet, from an educators' perspective, such games may still fall short with respect to the essential reflective component of learning, since immersive in-game learning may in general be too implicit, in particular when no clear criteria (*right* or *wrong*) exist, as for example in more complex management games, that involve training of communication skills, Eliens & Chang (2007).

In this paper, we will introduce (the first steps towards) a framework for understanding game play based on symbolic machinima, that is the recording of user actions such that later replay is possible, either in a realistic format as in 'normal' machinima or in an adapted format, symbolically, to allow for reviewers comments or even the revision of players decisions. In summary, we envisage a number of possible application contexts for our approach:

- find **exploration** and **learning** strategies in serious games – and to allow feedback and review by human experts in a **replay, think & reflect** context
- provide **user-tailored services** – e.g. museum guide, or personalisation and recommendation scenarios
- learn user's **multimodal behavior** – for control or as sample for NPCs (Non-Player Characters)

Although we are in the process of implementing our approach in the XIMPEL platform[1], we see our contribution on a high and general level, which may be applied for a variety of different platforms. In our own research, not only do we wish to use the developed technology in the context of serious games based on interactive video, but also in explorative educational environments for (virtual) musea. Another application domain is training (such as a virtual fitness trainer), where feedback strategies are of major interest, Ruttkay et al. (2006).

structure The structure of this paper is as follows. First we will briefly characterize serious games, in particular in distinction to e-learning and online information facilities. Then we will discuss the background and inspiration of our approach and characterize the notion of symbolic machinima. After dealing with issues in tracking user actions in virtual environments, we sketch the outlines of a behavioral model for game play, and discuss its relation to existing game reference models. We then define a metrics for behavioral discrepancy, which allows for relating user behavior to possible norm scenarios, and conclude by giving a brief overview of the benefits of our approach with respect to understanding individual users' game play as well as the use of analysis data for authoring alternative game scenarios.

[1] ximpel.net

SERIOUS GAME(S) RE-CONSIDERED

The literature on (serious) games abounds. To set the stage, we limit ourselves to a single quote from *virtual heroes*[2], that we more extensively discussed in Eliens & Chang (2007):

> *Serious games and simulations are poised for a second revolution. Today's children, our work-force and scientists are increasingly playing, learning, and inventing in visually intensive "virtual" environments. In our increasingly experiential economy, immersive educational and training solutions are needed to advance the workforce of tomorrow. Game-based learning and technologies meet this challenge.*

Yet, as we indicated in Eliens & Chang (2007), from the observation that serious game technology meets current educational challenges, it is still a long way to actually develop interesting serious games, that can not only compete with e-learning facilities in addressing educational goals, but may also be considered to be sufficiently playful to count as game(s) whatsoever.

A very helpful set of criteria for distinguishing games from other (online) applications were presented to us in a workshop on educational games in a museum context[3], which mentions 4 essential characteristics to assess the extent whether an application may considered to be a game:

- *challenge(s)* – relevance, feedback, confidence
- *curiosity* – cognitive or sensitive discrepancy
- *control* – contingency, choice, power
- *context* – intrinsic or extrinsic metaphor(s)

Leaving a more detailed interpretation of the characteristics *challenge*, *curiosity* and *control* to the inventiveness of the user, who may use the keywords for support, the *context* characteristic, however, needs some elaboration. As the discussion in the workshop indicated, it is rather easy to use extrinsic metaphors or game formats for arbitrary content. For example, a memory game can be reused over and over again, just by changing the images according to the topic or subject, that is language learning, climate change, etcetera. These kind of mini-games or casual games lend themselves to a variety of learning tasks and may be constructed using pre-defined game formats. More difficult, however, is to construct games with an intrinsic relation to the topic, and in the workshop for educational games in a museum the best suggestion was a scenario that imprisoned the player in the museum by night, haunted by the figures depicted on the paintings ...

[2] www.virtualheroes.com

[3] www.archimuse.com/mw2008/abstracts/prg_335001733.html

Apart from *context*, *challenge* and *control* seem to be the major parameters for modelling user actions in terms of, respectively, goals or topics and strategy and choices, or in other words scenarios with more or less well-established courses of behavior.

REALIZING SYMBOLIC MA-CHINIMA

The background and (partial) inspiration of our work is formed by the FP7-ICT-2007-3 Project CAESAR (Computer Aided Experiential Story Acquisition and Reuse), initiated by Pedro Gonzalez of the University of Madrid. Key elements in the proposal are *semantic-enabled machinima* and *end-user narrative content creation*. Not wishing to judge the wisdom of those who had to review the manifold of proposals, and notwithstanding the necessarily heterogeneous content of such proposals, imposed by the constraints of enforced cooperation between widely different international groups, we are nevertheless disappointed by the rejection of our request for funding, and decided to continue this valuable line of research within the proximity of our own groups.

Paraphrasing the CAESAR project proposal, we may observe that: The idea of producing animation movies using the tools and resources available in a game and rendering them with a 3D game engine, appeared in the early 90's and is now known under the term Machinima (machine cinema). Interest in machinima is growing as demonstrated by sites such as Tube2SL (tube2sl.com), a Second Life based Broadcasting Network for machinima productions, or WeGame (wegame.com), a media sharing platform for gamers, in public beta since January 2008, where gamers can post and share their in game recordings. New top of the line commercial games include machinima tools, such as Halo 3, Microsoft's XBox 360 flagship game title released at the end of 2007, whose Saved Films feature is one of the main innovations with respect to Halo 2.

Obviously there are many benefits to be expected when extrapolating machinima to semantic-enabled or symbolic machinima. Quoting the CEASAR proposal:

- **content creation** – With the rise of persistent worlds, content designers for video games are more in demand than ever. Roleplaying games have always needed truckloads of content, and now that these worlds have been made seemingly endless, the need for more content is ongoing. Game content generation by example would provide a highly cost-effective solution to this problem.
- **community building** – Community building is a key issue for games and even more for the success of virtual worlds. CAESAR would provide new and more powerful ways for the community members to share their experiences in th virtual world in a way that is not possible in the real world, exploiting full observability

and the possibility of living virtual experiences wearing somebody else's shoes.

- **demonstration material** – Animation movies have applications for entertainment, communication and education but their production is too expensive for most potential users. Semantic enabled machinima would open the range of users of animation movies with affordable cost.

Our specific role as partner in the CAESAR project was to address the issue of learning about user behaviour, in order to provide sensible feedback, from semantically logged interactions. Tracking representations are envisaged at a high level for which we would like to develop a general framework, which can be used as a reference for developing concrete methods for specific (replay/feedback/authoring) purposes in dedicated application contexts. All what is assumed is that user and system actions are logged with time stamps and semantic labels of some kind. Below we introduce the different aspects of the analysis and learning framework, as it pertains to

level(s) of behaviour

- **bodily aspects** – to learn communicative multimodal behaviour patterns, or new gestures or motion sequences. In this case the tags are low-level bodily signals (blink, nod, beat), which may be gathered by some sensory input devices (vision, pressure sensor).
- **cognitive aspects** – to analyse or learn strategies, action patterns of the user. The labels are on a high semantical level, indicating choices made by the user in given situations.
- **narrative/rethoric aspects** – analyse dramatic effects, features about the emerging narrative

USER TRACKING IN VIRTUAL ENVIRONMENTS

Many of the technologies to realize games or rich-media interactive applications, including X3D/VRML, Flex/AS3, and the Halflife 2 SDK, as well as Second Life, use events to capture user actions, which in its turn may be stored and programmatically invoked to re-create or simulate a sequence of user actions. In Eliens (2000), moreover, we have demonstrated how to use object-technology to create event-driven simulations capturing complex state information, allowing for complete undo and redo actions.
As reported in Eliens et al. (2007a), we used event-capture techniques to create guided tours in virtual environments for cultural heritage by tracking expert behavior, even allowing for the user of guided tours to express preferences for particular choices by (implicitly) defining weights on the influence of experts deciding on alternative choices. Thus having a database of tours from a number of experts, we may offer the user a choice of tours, and even allow to give priority to one or more of his/her favorite experts, again simply by adjusting the weighting scheme.
As more fully explained in Eliens & Wang (2007), guided tours, in the digital dossier, may take one of the following forms:

- automated (viewpoint) navigation in virtual space,
- an animation explaining, for example, the construction of an artwork, or
- the (narrative) presentation of a sequence of concept nodes.

In practice, a guided tour may be constructed as a combination of these elements, interweaving, for example, the explanation of concepts, or biographic material of the artist, with the demonstration of the positioning of an artwork in an exhibition space.
As a pre-condition for the construction of guided tours based on user tracking we identified the requirement that navigation consists of a small number of discrete steps. This excludes, at first sight, the construction of arbitrary guided tours in virtual space, since it is not immediately obvious how navigation in virtual space may be properly discretized. As an additional requirement, it must be possible to normalize interaction sequences, to eliminate the influence of short-cuts, and to allow for comparison between a collection of recordings. The application of the techniques developed for constructing guided tours requires that choices are discrete and only apply to capture navigation in virtual environments when we find find proper ways to encode such navigation as a small finite collection of discrete steps. Also in the discrete case, however, we must be able to normalize navigation paths, in order to compare and weigh the navigation sequences of multiple users.
For the actual playback, as a guided tour or replay, a decision mechanism may be needed that finds the proper advice or weight at each decision point to select the optimal step, according to some decision rule that takes the weighting scheme as for example expressed in a norm-scenario into account.
In Eliens et al. (2007c), we have indicated how tracking user behavior may be realized in Second Life using an elementary web-server containing the following resources:

web server

- /seen?user=SomeAvatar – records the presence of SomeAvatar
- /touched?user=SomeAvatar – invokes object API for user SomeAvatar
- /set_tag?user=SomeAvatar&tag=FavoriteTag – records SomeAvatar's favourite tag

For example, in response to a 'touch' event, invoking *touch* results in consulting the database for the user's tag and possibly sending a request to the object API performing some action on behalf of the user or recording a user's favorite tag. These invocations could easily be extended with time tags to enforce linear ordering.

TOWARDS A BEHAVIORAL MODEL FOR GAMEPLAY

Storing events resulting from user actions, possibly together with events influencing the game state autonomously, generated by the game system, gives us an immediate, albeit low level, way to record game play, allowing for machinima-like replay. However, in order to be able to provide meaningful feedback on the choices made by the user during play, we need a more high level representation of the users' behavior and choices made when the user is confronted with particular challenges. For inspiration, we first looked at what *game interaction patterns* might have to offer, Björk & Holopainen (2005), but apart from the *Score* pattern, little support was found for symbolically representing user actions, due to the rather abstract nature of patterns.

To simplify matters, we decided to reduce the representation problem to modelling the behavior of users at choice points in interactive video, as supported by the XIMPEL platform, and how particular choices reflect the attitudes or preferences of users with regard to particular topics. Although limiting ourselves to interactive video may seem to be too restrictive, as we argue in Eliens et al. (2008) interactive video may provide an excellent basis for game play, and as for example demonstrated in our Dante-inspired *Journey to Hell*[4] application, allows for assessing what we may call in this case a *moral profile* of the user, simply by recording the choices made by the user on questions of a moral nature. In the hope of being able to extend the model to more rich forms of game play, this approach allows us to take a model originally meant to capture ratings and recommendations, as explored in Eliens & Wang (2007), and extend this to represent attitudes and preferences with regard to topics of interest. See also Van Setten (2005) for information on recommendation and user modeling.

Since XIMPEL was originally developed for a climate game, Eliens et al. (2007b), we will take climate issues and attitudes towards measures affecting global warming or the effects thereof as a starting point to illustrate our approach, which we will present without going into very much formal detail. As an example, let's look at how we may model the behavior B of a user in the context C of a debate between experts, where the user is challenged to take action to provent flooding of the Netherlands due to global warming, for example by reducing the emission of CO2. In outline we may represent this situation as:

behavior = [choice = measures, action = reduction]
context = [context = debate, challenge = flooding]
preference(s) = [control(human_influence) = true]

[4] www.cs.vu.nl/~eliens/ximpel/dante

Here we represent *behavior* by relating actions to choice points, *context* by making the situation explicit in which the choice is presented as well as the challenge the user is confronted with, and finally *preferences* by indicatiog how the user takes control. Admittedly, for the derivation of preferences based on behavior in context we would need a rather strong ontology describing the semantic relations within the game domain. Nevertheless, although still a far cry from a formal model, having a suitable representation for *choices*, *actions* as well as the features defining *context*, *challenges* and *preferences* would allow us to record game events on a sufficiently high level, so that they may later be used for meaningful feedback.

A REFERENCE MODEL FOR EFFECTIVE GAME PLAY

In Eliens & Chang (2007) we introduced a reference model for game play, to be able to decide on the effectiveness of the players' strategies and actions in attaining the goals set in service managements games. The basic model, adapted from Juul (2005), consisted of the following elements:

reference model

- *rules* – service management protocols
- *outcome* – learning process
- *value* – intellectual satisfaction
- *effort* – study procedures
- *attachment* – corporate identity
- *consequences* – job qualification

Relating this model to our *challenge*, *curiosity*, *control* and *context* criteria, we may regard *rules* and *effort* as constitutive factors for *challenge*, *outcome* and *effort* as belonging to *control*, and *attachment* and *consequences* as belonging to *context*, to which, naturally, also *rules* bear a strong relation.

For service management games, we added two more criteria to the model, namely *scenarios* and *reward*, dealing with the (serious) content of the game:

- *scenarios* - problem solving service management
- *reward* - service level agreement

Both the notions of *scenario* and *reward* are essential in understanding (serious) game play, since they allow to indicate a specific level of attainment to which the player must comply, in order to be considered to have played the game effectively.

From a different perspective, in terms of the notions introduced in a behavioral model of game play, as introduced in the previous section, we may classify *rules* as belonging to *context*, *outcome*, *consequence* and *value* to *preferences*, and *effort* to *behavior*. Adding *scenarios* and *rewards* helps in defining *challenges*, and, in principle, to define *norm scenarios*, setting a standard for the most appropriate actions, which is clearly relevant for serious games intending to bring about an attitude change, for example in behavior affecting climate change.

METRICS FOR BEHAVIORAL DISCREPANCY

Given the notions of *scenario* and *reward*, as introduced in the previous section, we cannot resist to speculate on how we can define *norm* scenario(s) and associated metrics to assess behavioral discrepancy, that is the degree in which the user deviates from a desired course of action, a particular position or set of preferences. To allow for a more formal treatment, it seems most convenient to adapt the behavioral model introduced earlier, by reducing behaviors to consist of actions only, taking context into account implicitly, and to redefine preferences as rewards, which may conveniently be expressed as scores over predefined result parameters, such as $(people, planet, profit)$ in the case of our climate game.

Representing the combined result parameters $(people, planet, profit)$ as vectors of features characterizing preferences for aspects of the individual parameters allows for defining a metric over the space of preferences defined by the result parameters, using a standard distance metric, as we originally did for recommendations in Eliens & Wang (2007).

Using such a metric allows us to assign a rating or an indication of relevance to the result parameters, as illustrated by the following example. If we assume that alternative actions have effects as listed below

a_1 = [planet = green, profit = high]
a_2 = [planet = green, people = happy]
a_3 = [planet = red, profit = high]
a_4 = [planet = red, people = happy]

then we may, in an abstract fashion, deduce that if $d(a_1, a_2) < d(a_1, a_3)$ then $r(profit) < r(planet)$, for a rating funtion r.

However, if $d(a_1, a_3) < d(a_1, a_2)$ the reverse is true, that is then $r(planet) < r(profit)$. In other words, actions involving only particular features of any of the result parameters may influence the final result when taking a particular position or preference as the norm.

Given a metric on preferences, which induces a metric on actions, and a norm scneario, with a recommended sequence of situations $\ldots, s_{n-1}, s_n, \ldots$, with for s_n possibly alternative actions $a_1, a_2, \ldots$, we may adapt the (implied) preference of the user, when the user chooses to select alternative a_k instead of accepting s_n as recommended by the norm scenario, to adjust the score by taking into account an additional constraint on the derived score. Differently put, when we denote by $s_{n-1} \mapsto s_n/[a_1, a_2, \ldots]$ the presentation of issue s_n with as possible alternative actions $a_1, a_2, \ldots$, we know that $d(s_{n-1}, a_k) < d(s_{n-1}, s_n)$ for some k, if the user chooses for a_k.

Admittedly, apart from easily skipping over representational issues, we have omitted many of the necessary formal details. We refer to Cesa-Bianchi and Lugosi (2006) for readers wishing to explore the mathematical details of our approach.

BENEFITS OF A QUALITATIVE APPROACH TO REFLECTION AND FEEDBACK

Whereas quantitative results, as for example obtained in tests or exercises in specific skills, may be worthwhile in domains such as language learning or, for that matter, the operation of vehicles, a more qualitative approach seems necessary for (serious) game tasks that involve communication skills or strive to induce attitude changes, as is the case with management games or games related to topics of societal interest, such as climate change and security.

With respect to individual users, an approach as sketched in this paper offers the opportunity to analyse behavioral patterns of a single user interacting with the system. The issues involved here, or the potential usage of such analysis include:

- one particular user vs. group behavior
- one session or multiple sessions
- novice vs. experienced user or expert
- possible recommendations or advice

Our approach, which we have summarized in the title of this paper as *record, replay & reflect* originated from the wish to provide feedback and replay, preferably in a user-friendly textual format, that is to present segments of the interaction for viewing interesting/problematic parts, give summary about the interaction and performance (e.g. in a learning environment), either for the user him/herself or as a summary for system developers providing feedback about the usage of the facilities within the system. Other goals for which our approach may be used, as expressed in the original CAESAR proposal, include:

goal(s) of analysis

- enhance the **behaviour repertoire** of user-control or of virtual characters – in the game, based on analysis of user behaviour and interpreting it on a higher level (e.g. new types of greeting signals, new 'escape' behaviour).

- author exploration paths and navigation strategies – use the interaction sample (possibly by an expert) as a reference e.g. for selective tours in a virtual environment or learning applications.
- learn user profile(s) – elicit typical interaction or behaviour characteristics of a given user or user group, and use this information to re-design or dynamically adapt the system.

Finally, as also mentioned in the CAESAR proposal, where the deployment of semantic-enabled machina for *content authoring* played a central role, we envisage the potential reuse of game play in different contexts or platforms. With a sufficiently high level representation in a suitable interchange format such as XML, we would also like to explore the reuse of missions and scenarios in different contexts, and even different platforms, similar as the proposed Collada[5] standard does for (graphics and physics) game content, as a means to accommodate the authoring of narratives and story lines.

CONCLUSIONS

In this paper we have sketched the outline of a framework for understanding game play, which may be used for providing meaningful feedback to (serious) game players, allowing for replay on a sufficiently abstract level by deplaying semantic-enabled machinima.
Although we have partially implemented aspects of our approach in the XIMPEL platform, addititional research is needed to arrive at a sufficiently complete representation scheme for capturing events, user actions and resulting game state changes. It seems that, in particular, we must pay more attention to the domain ontology underlying a game, to enable the construction of user profiles using inferential reasoning based on the actions taken by the player when confronted with choices or challenges in the pursuit of a mission or scenario.

Acknowledgement(s) Thanks are due to the guys from the Clima Furura Labs, Marek van de Watering, Hugo Huurdeman and Winoe Bhikharie, who among other things developed the XIMPEL platform. We also wish to express our thanks to Pedro Gonzalez of the University of Madrid who initiated and got us involved in the CAESAR project.

REFERENCES

Björk S. and Holopainen J. (2005), *Patterns in Game Design*, Charles River Media

Cesa-Bianchi N. and Lugosi G. (2006), *Prediction, Learning, and Games*, Cambridge University Press

[5]en.wikipedia.org/wiki/COLLADA

Eliens A. (2000), *Principles of Object-Oriented Software Development*, Addison-Wesley Longman, 2nd edn.

Eliens A., Wang Y. van Riel C. and Scholte T. (2007), 3D Digital Dossiers – a new way of presenting cultural heritage on the Web, In Proc. Web3D 2007, ACM SIGGRAPH, pp. 157-160

Eliens A., van de Watering M., Huurdeman H., Bhikharie S.V., Lemmers H., Vellinga P. (2007), Clima Futura @ VU – communicating (unconvenient) science, In Proc. GAME-ON 07, Nov 20-22, University of Bologna, Marco Roccetti (ed.), pp. 125-129

Eliens A. Feldberg F., Konijn E., Compter E. (2007) , Mashups in Second Life @ VU, In Proc. GAME-ON 07, Nov 20-22, University of Bologna, Marco Roccetti (ed.), pp. 130-134

Eliens A. and Chang T. (2007), *Let's be serious – ICT is not a (simple) game*, In Proc. *FUBUTEC 2007*, April 2007, Delft

Eliens A. and Wang Y. (2007), Expert advice and regret for serial recommenders, In Proc. EUROMEDIA 2007, L. Rothkrantz and Ch. van der Mast (eds), Eurosis, pp. 111-118

Eliens A., Huurdeman H., van de Watering M., Bhikharie S.V. (2008), XIMPEL Interactive Video – between narrative(s) and game play, In Proc. GAME-ON 2008, Valencia, Spain

Juul J. (2005), *Half Real – Video Games between Real Rules and Fictional Worlds*, MIT Press

Gee J.P. (2003), *What video games have to teach us about learning and literacy*, Palgrave Macmillan

Ruttkay Zs., Zwiers J., van Welbergen H. and Reidsma D., Towards a Reactive Virtual Trainer, Proc. of IVA 2006, LNAI 4133, pp. 292-303. 2006.

Van Setten M. (2005), *Supporting People in Finding Information – Hybrid recommender Systems and Goal-based Structuring*, Ph.D. Thesis, Telematica Institute Netherlands

Anton Eliens (PhD) is lecturer and coordinator of multimedia at the VU University Amsterdam, and was recently appointed at the University Twente as professor creative technology / new media. He has been closely collaborating with Zsófia Ruttkay and is experienced in web-based interactive media such as Second Life, interactive video, and the application of such technologies in serious games like Clima Futura.

Zsófia Ruttkay (PhD) is Assoc. Prof. at the University of Twente and leads the Creative Technology Working Group. She has expertise is in creating styled multimodal behaviour and communication strategies for virtual humans. She also has a strong interest in educational games, and has been developing an interactive virtual trainer application.