Narrative in Virtual Environments - Towards Emergent Narrative

Ruth Aylett

Centre for Virtual Environments
Business House, University of Salford
Salford M5 4WT, UK
r.s.aylett@salford.ac.uk

Abstract

In this paper we consider the clash between the pre-scripted character of much narrative and the freedom afforded by a Virtual Environment. We discuss the concept of *emergent narrative* as a possible way of avoiding this clash. We examine the role of a VE user in a narrative and consider the concept of *social presence* as a means of reconciling the freedom of the user with the constraints of an emergent narrative.

Introduction

This short paper describes preliminary thoughts - rather than completed work - on narrative issues arising from a recent project 'Virtual Teletubbies' [Aylett et al 99]. In this project, children's TV characters - Teletubbies - were incorporated into a virtual environment (VE), by which we mean a 3D graphically rendered world in which both they and the user have a joint spatial existence. In this case, the user was not required to use special immersive hardware, such as a head-mounted display, but was represented by an invisible camera position attached to spatial controls which allowed him or her to 'move' within the VE.

The Teletubbies were implemented using a behavioural robot architecture, in which emergent behaviour at a given moment is determined by the synthesis of responses from currently active behaviour patterns [Barnes 96]. These behaviour patterns are in turn driven by simple virtual sensors while groups of behaviour patterns (packets) are activated or deactivated according to the level of the Teletubbies' internal drives, such as hunger, fatigue and curiosity.

As with robots, the initial behaviours implemented were those allowing physical movement in the environment, in this case taking into account the sloping nature of the outdoor terrain being modelled. A simple form of gravity was also modelled so that neither Teletubbies nor user were able to 'fly' within the VE. Virtual sensors were implemented as bounding boxes and as a forward sweeping sensor for obstacle avoidance; information about objects in a Teletubby line-of-sight was transferred directly as symbolic information rather than through a virtual retina.

While this project was successful in examining some basic architectural issues, a number of issues of narrative arose, many of which are common to virtual environments of other types. We consider the problem of narrative as it arose in this particular project, examine alternative ways of dealing with the issues, and sketch out a way forward which involves an approach to narrative we believe has not yet been very widely investigated.

A Narrative Problem

In moving characters from television to VE, the difficulty in retaining the original narrative approach was immediately evident. On television, the narrative is normally wholly pre-scripted and is viewed passively by the audience from camera positions determined by the creators of the narrative. However, translating this directly into a VE removes all the characteristics that differentiate a VE from television. The ability of the user to interact individually with the VE and even their ability to select their spatial viewpoint is denied in this approach.

While the narrative *experience* fails to exploit the characteristics of a VE, there can of course be a change in the process of narrative *production*. Through the provision of an appropriate toolkit, characters in a VE can be treated as virtual actors and the user can construct narratives in which they appear. This approach has been investigated by a number of groups, in 2D multi-media environments as well as in VEs, as for example the 'Virtual Theater' project of Barbara Hayes-Roth [Hayes-Roth & Brownston 95] and the IMPROV virtual actors system [Goldberg 97]. As well as empowering users who would normally be excluded from the creation of 3D spatially located narrative, this approach is also of interest to the film industry. Directors are beginning to exploit the power of 3D graphic environments explicitly in films such as 'Toy Story' and 'Antz' and implicitly via special effects in films such as 'Titanic'. Nevertheless, the extension of freedom in the process of narrative construction does not in our view remove the need to examine how the narrative *experience* might be extended to take advantage of the specific characteristics of a VE.

Two issues are of interest here. The first is how far the pre-determined nature of much narrative can be relaxed. The second is how far the user of a VE can freely participate in a narrative rather than acting as a spectator. Clearly these issues are related - a wholly pre-determined narrative also pre-determines the role, if any, of the user within it. Thus relaxing this constraint allows more freedom for user

involvement. However it does not of itself settle the issue of exactly how the user can be incorporated into a narrative.

Narrative approaches

One way of examining narrative approaches is to divide the narrative process into a number of levels - a not uncommon technique in robotic and agent architectures - and to consider them top-down. Without making any claim for a definitive hierarchy here, one can produce something like this:

Overall plot

Character-level abstract action sequences

Physical behaviour   - cognitively determined

                           - reactively determined

Thus a narrative may have a plot in which boy meets girl, boy loses girl, boy performs heroic feat, boy regains girl. This plot can be achieved by a number of different abstract action sequences. For example the first element could be implemented as: come into the room, walk up and say hello; creep up behind the character and say 'boo'; stand close to the character near others and join in an existing conversation; and so on. In turn each such action can be embodied in different physical behaviour. This physical behaviour can be thought of as more cognitively determined by the character: exactly what words to say, for example; or more reactively determined: as in stumbling or laughing involuntarily.

At one extreme, a narrative may be wholly scripted down to the most detailed level. For example, a film director often goes to great lengths and many takes to achieve exactly the desired effect with dialogue, body language and facial expression, not to mention every physical item in the setting. Deviations from this scheme can be excluded from the finished film. From this perspective, virtual actors who never tire of retakes and do exactly what they are programmed to do may be an improvement over the real thing. A somewhat looser position can be seen in live theatre, where for example a classic text may be reinterpreted via variation in the speaking of the text and in the setting. However in a VE, both correspond to pre-scripted animation in which the narrative is translated into virtual agent actions in exactly the same way every time the narrative is executed.

Further along the spectrum, both in filmed and live media, improvisation may be used. Here elements of the narrative are created dynamically, though often within overall constraints. As these constraints are set at higher levels of the hierarchy above, the improvisational element becomes greater. The abstract action sequences characters carry out may be fixed, but physical behaviour may be wholly improvised, so that the words used are dynamically generated. Alternatively, the overall narrative structure may be fixed at a general level but the exact actions characters carry out as well as the exact words they speak may be improvised.

Improvisation is investigated in VEs, but mainly at very low levels in the hierarchy above. A number of systems allow sensor-driven reactive behaviour by virtual agents. For example, training environments have been created in battlefield medical first-aid [Stansfield et al 98] in which a virtual patient's physical state changes according to the first-aid administered by the user, and in hostage release [Shawver 97] in which virtual hostages dive to the floor if the user, acting as rescuer, fires a shot. Improvisation at the level of cognitive behaviour is less developed, probably because it is mostly expressed via natural language and thus poses difficult problems. Virtual agent improvisation at the level of action sequences is also infrequent as this requires some form of planning, whether through a generative approach or plan selection from libraries.

Could one speak of improvisation at the topmost level of the hierarchy above, that of overall plot? This may appear to require formidable creative abilities, which are difficult for humans (storytellers, playwrights, novelists, film directors) and wishful thinking in the current state-of-the-art of intelligent agents. However, the ideas behind the behavioural approach used in the Teletubbies project suggest an alternative view. The principle followed by behavioural architectures is that of emergence: the creation of complexity bottom up via interaction between essentially simple components. Rather than seeing the hierarchy above as a top-down structure, one may view it as a bottom-up structure, in which each level is created by interaction below it.

Emergent Narrative

Emergent narrative may seem paradoxical since the underlying structure provided by a definite plot (or equivalent - a plot in the classic sense may not be the only type of high-level narrative structure) seems needed to make a narrative 'hang together'. Yet in an obvious sense, narrative is emergent, since it has emerged from human life experience. If 'life' seems too grandiose a concept to fit into a VE, there are smaller-scale examples in which explicit narrative structure is absent but narrative frequently emerges through interaction.

Team games form a particularly obvious example, as in the case of football (as a UK group we mean soccer here). At the level of physical behaviour, football is indeed a group of people (usually men) kicking a ball about. The addition of conflicting aims and some constraints on allowable physical behaviour together with a limited time often - though by no means invariably - produces recognisable narrative structure. For example, the late substitution producing the winning goal; the new young player scoring on his debut; the player committing a reprehensible foul who injures himself seriously in the process; the talented but petulant player who retaliates when fouled, gets sent off, and loses his team a crucial match. Arguably it is this emergent narrative structure that gives football and other multi-player games an appeal different from say gymnastics.

In order to experiment with the concept of emergent narrative in a VE, for example with the Virtual Teletubbies, a number of issues must be investigated. What structures are needed to produce narrative often enough and with enough complexity to satisfy the user? There are a lot of very boring football matches too. In the 1980s, UK television carried a live broadcast of an embassy siege being broken up by

commandos. This consisted of nothing visible happening for some time and then a series of fairly mysterious events happening very quickly indeed. The narrative reconstructed after the event with suitably edited material had far greater coherence and power in this viewer's estimation.

A useful area for investigation is that of 'free improvisation' in drama. Arguably the basis for a 'free improvisation' is also the basis for an emergent narrative, since actors will normally try to produce some kind of emergent narrative framework in order to hold the attention of an audience. This type of improvisation seems to require actors to have established characters and some kind of relationship between these characters. An overall goal or precipitating event is often specified, usually involving some kind of conflict between the characters. Thus: a teenage daughter in a very middle-class home tells her mother that she is pregnant; two teenage girls clash because one of them has snogged the other's boyfriend.

Comparing this with a football match we see that there too character is specified (being life, each footballer plays themselves) and relationships are also specified: both between teams and within teams. There is also an overall goal in both senses of the word. The hostage release incident, on the other hand, had no characterisation of the individuals involved, and no close relationship between them, at least from the television camera position (it may or may not have formed a more satisfactory narrative from the perspective of a hostage).

One may apply this to the Teletubbies. In fact, the narrative structure is usually rather simple, since the television programme is aimed at children aged two or even younger. So, for example, a small pink cloud invades the Teletubby world, and the Noo-noo (a vacuum cleaner on wheels that acts as an autonomous character) tries to clean it up. However as the Noo-noo is not designed to have clouds inside it, this produces some anomalous behaviour as it tries to clean up things which are not rubbish, like the Teletubby toast machine. Eventually it expels the cloud and returns to normal.

This seems a very plausible emergent narrative assuming a sufficiently rich set of Noo-noo behaviours - in this case both internal and external. The precipitating event would be the arrival of the cloud near the Noo-noo. This suggests that in general a suitable number of dynamic events outside the control of the characters are needed in order to produce interesting behavioural reactions. For example, the Teletubby curiosity drive is more likely to form the basis for emergent narrative if unfamiliar objects appear in their environment.

We have commented that 'a sufficiently rich' set of behaviours is required for the characters in an emergent narrative without defining what this is. Experiment is one way of exploring this issue. However it seems likely that one avenue that should be incorporated in the Teletubbies is a more sophisticated emotional system than the basic drives that Teletubbies currently possess. These drives are already used to produce sequences of Teletubby action: a hungry Teletubby will head for the dome to get some food; curiosity, as just commented, will lead a Teletubby to investigate novel objects. A wider range of emotions connected into the architecture in a more principled manner would support more interesting interaction between Teletubbies and other elements of the environment.

We have already commented that one of the risks of emergent narrative is that it may not emerge - the unpredictability that makes it interesting also makes it in some sense fragile. One might add that it is also inherently small-scale: like free improvisation it runs in continuous time, there is no 'leaving out the boring bits' as would be the case in a written narrative. Real drama deals with this problem by breaking its narrative into scenes, so that activity the dramatist does not want to present to the audience can occur in between. One could of course implement this in a VE - flipping the environment onto 'the next day' for example. However since episodic drama often requires that characters to behave as if they have experienced the 'between scenes' activity, one might have to compute this for the virtual agents.

The role of the user

A second set of issues to be confronted in experimenting with emergent narrative surrounds the role of the user, not so far considered. We have argued above that the freedom of the user to interact with the VE and in particular to 'move' freely within it are essential features - indeed it is this which contributes so greatly to the feeling of *presence* for which many VEs aim. However it does pose problems in establishing a user role within a narrative structure. The comments on the hostage release example above underline that a narrative may only emerge from a particular perspective and may fail entirely if the user is in the 'wrong' position. We should note that even the role of spectator in a pre-scripted narrative becomes problematic in a VE, since the user may not be visually present in the position that was assumed when the narrative was constructed.

This issue has been confronted in the domain of computer games, but the solutions adopted there seem of limited applicability since they rely on the user constructing the abstract action sequence as the central active character. In adventure games, such as Myst, there is often an enveloping narrative, but in reality there is little organic connection between this and the action sequence created by the user. The abstract action sequence is often sketchily - if at all - realised in physical behaviour. Other characters appear rarely and do not construct any developing relationship with the user that would help to anchor the user's actions to the overall narrative.

On the other hand, in arcade-type games, physical behaviour is dominant with little sense of abstract action sequence or overall plot other than 'kill the monsters' and/or 'collect the treasure'. In both types of game, the ability of the user to 'undo' any amount of behaviour without consequences undermines the concept of narrative coherence. The user may have as little idea by the end of the game of his or her narrative path as a film actor who has appeared in dozens of takes of scenes in various orders but has not seen the final film.

The concept of emergent narrative suggests a different approach to the user's role. Enough interaction between virtual characters must exist, independently of the user's role, to produce narrative in this way. By analogy with emergent behaviour at lower levels, narrative will only emerge through the right type of components interacting in the right type of way. If the user is to contribute, they have to provide the appropriate behaviour. It may be difficult even for a co-operative user to be sure what behaviour is appropriate in this sense - after all, it is the high chance of 'getting it wrong' which lies behind the undo facilities in games.

Various mechanisms exist for helping a user to contribute positively to an emerging narrative. A simple one is the kind of pre-briefing favoured both in adventure game booklets and in role-playing games used in training. This is a straightforward way of postulating a shared history between the user and the environment and a set of role-specific behaviours.

However a more interesting, though not mutually exclusive approach, extends the idea of presence referred to above. Presence in its classical sense is very much physical presence - the illusion of being physically located in a VE. In the real world, appropriate behaviour is very much socially determined. At a simple level, if a football player violates the physical constraints of a game, he is sent off the pitch by the referee. Moreover, the crowd may well express strong and noisy disapproval. We could term the creation of this kind of social pressure in a VE 'social presence'.

Social presence is therefore an extension of physical presence - the illusion not only of physical location but also of social location. If one could achieve this, then the behaviour required for emergent narrative could be communicated through social convention or social pressure. The large-scale MUDs and MOOs suggest that social presence can be produced with very limited amounts of physical presence, though it is worth remembering that these environments are almost entirely populated by human users with a full range of natural language and sophisticated social and cultural assumptions.

Conclusion

The Virtual Teletubbies domain has some advantages as one in which to explore emergent narrative and social presence. The complexity of narrative required is rather low if one uses the narratives produced in the TV programme as a yardstick. Moreover, as a VE aimed at young children, the social environment might also be adjudged simple. Young children often meet and play with others at nursery school or in parks without needing a previous history of social interaction within which to locate this.

On the negative side, there are obvious difficulties in trying out such a VE on an appropriate set of users. Not only are children of the right age hard to access, there are obvious problems in creating an interface which will allow natural interaction. A head-mounted display seems out of the question, and it could be that a CAVE or similar interface

would really be needed to avoid the interface itself becoming a major barrier to the VE.

In conclusion, we argue that the combination of emergent narrative with the creation of social presence provides a little explored avenue for adapting narrative to the particular characteristics of VEs.

References

References

Aylett, R.S; Horrobin, A; O'Hare, J.J; Osman, A. & Polyak, M. (1999) "Virtual teletubbies: reapplying a robot architecture to virtual agents" Proceedings, 3rd International Conference on Autonomous Agents (to appear).

Barnes, D.P. (1996) A Behavioural Synthesis Architecture for the Control of Mobile Robots. In: Advanced Robotics and Intelligent Machines, ed. J.O.Gray & D.G.Caldwell. IEE Control Eng. Series 51, IEE London 1996

Goldberg,A.(1997) "Improv: A system for real-time animation of behavior-based interactive synthetic actors". In R. Trappl and P. Petta, editors, Creating Personalities for Synthetic Actors, pages 58-73. Springer-Verlag, 1997.

Hayes-Roth, B. & Brownston, L. (1995) "Multiagent collaboration in directed improvisation". In Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95), pages 148-154, San Francisco, CA, June 1995.

Shawver, D. (1997)"Virtual Actors and Avatars in a Flexible, User-Determined-Scenario Environment," Proceedings of the Virtual Reality Annual International Symposium, Albuquerque, NM, 1-5 March 1997

Stansfield, S; Shawver, D. & Sobel, A. (1998) "MediSim: A Prototype VR System for Training Medical First Responders," Proceedings of the Virtual Reality Annual International Symposium, Atlanta, GA, 14-18 March, 1998