

Interactive Storytelling in Virtual Environments: Building the “Holodeck”

Marc CAVAZZA (1), Ruth AYLETT (2) Kerstin DAUTENHAHN (3),
Clive FENCOTT (1) and Fred CHARLES (1)

(1) *University of Teesside, Middlesbrough, TS1 3BA, UK*
{m.o.cavazza, p.c.fencott, f.charles}@tees.ac.uk

(2) *University of Salford, Salford, M5 4WT, UK*
r.s.aylett@salford.ac.uk

(2) *University of Hertfordshire, Hatfield, AL10 9AB, UK*
k.dautenhahn@herts.ac.uk

Abstract. In this paper, we introduce an ongoing research project, which aims at developing an immersive storytelling environment. In such a system, which is inspired from the “Holodeck” device, human users will be able to interact with artificial actors through speech, while sharing the same physical environment. Once on the set, the users will be able to participate in story generation by influencing virtual actors or altering the common environment. An essential aspect is that user involvement will not have to be permanent, and that the users will be allowed to leave the set to watch the story unfold in their absence and to return at a later stage. We discuss some fundamental AI problems with which the implementation will be faced and relate them to narrative theories and models of storytelling. More specifically, we describe how users can interfere with the virtual actor’s plans to influence story generation in terms of interaction

1. Introduction

Our “Holodeck” project aims at developing an immersive story generation system supported by an Intelligent Virtual Environment. The system would appear to the user as an immersive virtual environment enabling physical interaction (i.e. through a CAVE or other immersive projection display). In this environment, the user would be able to interact with a cast of virtual actors (VA’s) sharing the same environment and addressing them in natural language. An essential feature is that the user involvement should not be a permanent one. Instead, the user should be allowed to participate in the action, helping to create the story and leave the stage to watch the story unfold as a consequence of his/her intervention. Another essential aspect is that the user would still be able to get back into the story at some suitable time, or even to influence it from his spectator position, for instance by interfering with the emotional state or the behaviour of VA’s. This new form of entertainment would implement in its own way the convergence of VR, cinema and computer games as described for instance by Altman and Nakatsu [1] (see Figure 1), and has obviously been influenced by [2]. The relations between VR and storytelling are complex and even though our main purpose is to develop an interactive storytelling system, these relations are worth exploring into some more details. Story-telling and narratives are more than an enjoyable form of entertainment. Humans are social animals, a child needs a social environment in order to grow up and become a social being who is able to survive in human society. It has been argued that narratives are the fundamental structure by which humans organise their experiences and

memories, in particular of the social world. New events are interpreted in terms of old stories, we know what we tell and we tell what we know (Schank & Abelson in [3]; [4]).

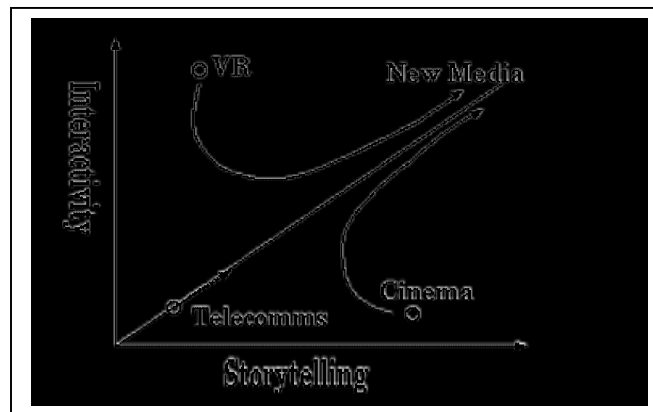


Figure 1. Convergence in VR and Storytelling (courtesy of Dr. Ryohei Nakatsu).

Story-telling plays an important role in the development of a child's understanding of the world and its relationship to the world [5]. From an evolutionary perspective, it is argued that stories have been the most efficient means of understanding and communicating about the social world (Read & Miller in [3]; [6]). Narratives are more than representations of reality, human knowledge about 'reality' is constructed around stories. On the level of the individual, stories are important in order to construct and re-construct one's autobiographical 'self', or 'centre of narrative gravity' [7]. According to Bruner [4] narrative operates as an instrument of mind in the construction of reality. Thus, story-telling in an intelligent virtual environment (IVE) is more than an entertaining add-on, it is a important requirement in order to create a believable (virtual) world, in order to allow a human to be socially immersed and being able to understand this world. Believable virtual environments need to be social environments in order to support a user's social presence [8]. An actor in the "Holodeck" scenario needs to be social, it needs to know about other agents and their relationships in the virtual environment, in particular it needs to know about the human. Additionally, a believable social world is a world which presents itself in terms of narrative, so that it can meet a user's cognitive needs to interpret, understand and interact with the world in terms of stories [9]. Basic social requirements of such narrative agents are 1) the ability to recognise and identify other agents, 2) the ability to establish and memorise direct (one-to-one) relationships with other agents, and 3) the ability to monitor and memorise third-party relationships, namely relationships between other agents in the shared environments [6].

Generally, story-telling and narrative is understood in term of human language. Here, virtual characters are able to interact with a human by means of spoken or written language (e.g. [10]. However, in addition to such 'stories-in-words' a narrative can also be expressed as 'stories-in-actions'. Thus, in the "Holodeck" scenario the believability of the narrative virtual actors will depend on their ability to communicate with each other and the human by using language (in particular in situations where the human is a spectator or 'on-stage' with the other agents), but narrative can also be expressed in non-verbal behaviour and the way actions are organised and expressed in a consistent fashion [11].

Where narrative is created 'on the fly' bottom-up from interaction, as we are proposing, rather than as is more usually the case, top-down from a script, there may be no single narrative thread. Different narratives might be extracted from the same mesh of events and interactions from the perspective of different characters. The focus of this work is to create

an environment in which events and interactions allow the user to 'storify' his or her experience, whether derived from participation or from observation of other characters.

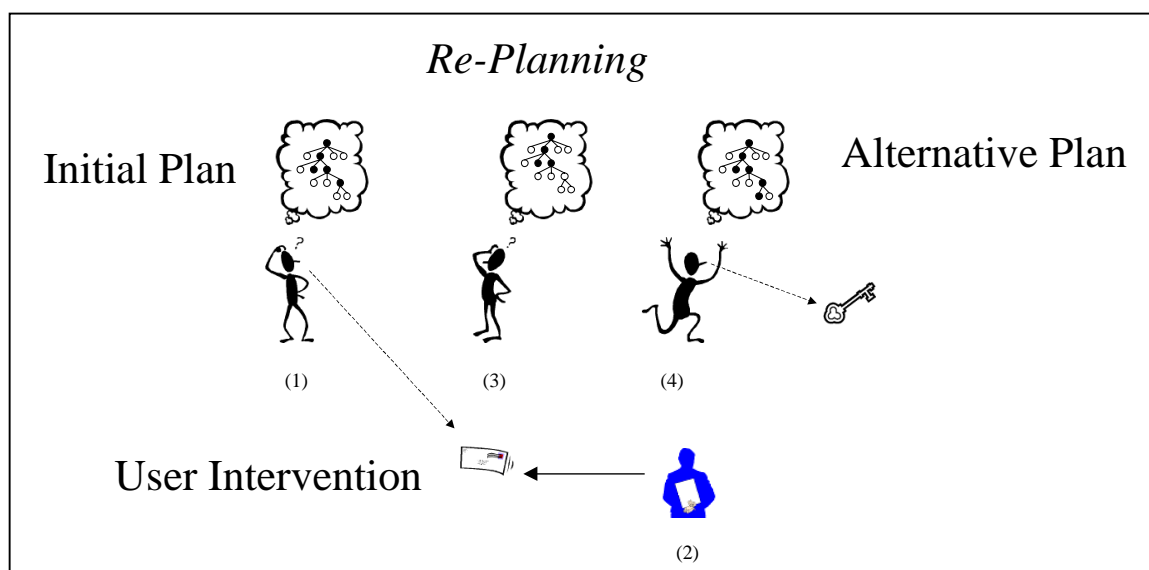


Figure 2. User Interfering with VA's Plans.

However, a tension exists between the privileges awarded to the user over other characters in order to support their ability to create a personal narrative, and the satisfaction they obtain from participation, which must to some extent depend on social immersion - 'believing' in the VAs and their world - and therefore in its autonomy. Thus the user's role and powers must be very carefully considered and are discussed below.

2. Interactive Story Generation

The ways in which the user can participate in story generation are i) through physical interactions with objects in the virtual environment ii) through on-stage conversation with artificial actors, and iii) through off-stage intervention (advice, interjections, etc.) What determines the typology of user intervention is a combination of causal input and interaction channels. Physical intervention, such as removing a critical object (handgun, letter, key...) from the set (Figure 2), requires the user to be part of the virtual stage. Interacting on the set with a character, like an actor would do, also requires physical presence on stage, though the input modality is speech rather than physical manipulation of objects. Conversely, influencing VA's in real-time ("take that letter", "don't open that box") can take place from outside the set.

Propagating the consequences of user intervention assumes some causal representations. Within a storytelling environment, there are further difficulties arising from the conjunction of different sorts of causality: i) physical causality, ii) psychological causality and iii) narrative causality. In the next section, we discuss each of these.

Physical causality is part of our common sense knowledge of the world. It is required to implement the causal consequences of physical manipulation of some (simple) systems and devices without having to actually simulate the physical processes in a numerical fashion. However, complex physical causality is not involved in most narratives (if one excepts specific cases, such as some sports performance being part of the plot at some stage), and might not be relevant in our context.

Psychological causality, whether a psychological fact or not, is generally taken for granted in storytelling. It relies on states of mind, which can be causally influenced by events, more

often by inter-agent communication. In AI terms, this corresponds to the fact that an agent's intentions are equated with its top-level goals [12]. It plays an essential role in relation to narrative functions (e.g. betrayal can be described as "supporting, then contrasting another actor's plan") and in having the character's psychology accessible to the spectator, which is part of the narrative experience.

Narrative causality is part of the reconstruction of the story from its events as they appear on stage. It can be seen as a redescription of events at a different level of granularity, leaving out irrelevant details and relating the main narrative functions one to another. It plays an important role in storytelling. As Raskin [13] suggests, "causality is necessary for the coherence and inner logic of a story, and when causality is conceived in terms of the power of characters to act on the situation in which they find themselves, it serves the interest of the story".

3. Underlying Narrative Theory

There exist many narrative theories since Aristotle's *Poetics*. The best known contemporary theories include Propp, Todorov, Greimas and Barthes. We base the early prototype of our system on Barthes' narratology, as described for instance in [14] and [15]. The rationale for adopting Barthes' theory is that its level of detail appears compatible with bottom-up story generation, which is required for an interactive system. Also, structuralist theories provide descriptions that can support formalisations, hence computer implementation (at least to some extent).

Barthes distinguishes several levels of complexity for narrative analysis. The level of narrative functions, the level of narrative actions and the story level, at which integration takes place. We will mainly make use of his notion of functions, which is a refinement of Propp's 31 narrative functions. Functions are elements of a correlation; they create the possibility for future events to happen and by doing so relate key actions to their consequences in the future. For instance, purchasing a gun can be related to the moment in the story where it might be used, with dramatic consequences. In this example, the validity of the "gun" function stems from the fact that using the gun as well as deciding not to use it can each be attributed meaningful interpretations for the story. Functions should thus be seen as elements bearing a strong meaning, from which branching points can be elaborated. Barthes calls these branching points dispatchers. Dispatchers are defined as the zones of risk in the plot, where causality is at stake. These need not be intrinsically dramatic, but can consist in events that take on a deferred signification, such as catching/missing a train, answering/not answering a phone call, opening/not opening a letter, etc. But we can already see that dispatchers are elements of content that should trigger expectations, hence some form of interpretation procedure to be applied on the causal chain of events. Hence an important part of storytelling, including automated storytelling, lies in the proper recognition and processing of dispatchers.

4. Mapping Causality to Narrative Structures

We can now investigate further the problem of narrative formalisation. There are two aspects to such a formalisation: i) the VA's behaviour should be supported by knowledge representations and ii) the events generated by the VA's should be recognised and integrated into a narrative structure, representing the story as an instantiation of the plot. These knowledge representations must obviously be fully integrated with the virtual environment.

Such an integration is facilitated by proper mechanisms for action simulation and action recognition [16].

The main driving force for the story consists in actions taken by its individual. A first requirement for a real actor improvising a drama - a close analogy to the process of bottom-up story generation - is to 'stay in role'. That is to say that the actions carried out should be 'in character' - appropriate to the persona, and also to the social constraints derived from the scenario. Thus causality must be filtered through the supposed history, emotional make-up and social role of the VA. These are partially expressed through motivations - the internal needs of a VA and the goals derived from them, and the plans that a VA therefore makes. However, the classical 'rational individual' who is wholly guided by their plans is not only a poor match with the real world, it is also unlikely to produce interesting stories. Nor would it support the believability of the VAs for the user, a vital requirement for successful user participation and for social immersion. Interaction and reaction are a fertile source of dramatic interest - the change in emotional states between two VA's during the course of an interaction is in itself a narrative. Reactions may also interact with motivations and plans in interesting ways - a VA with a bad temper may therefore shout at another VA whose co-operation is needed for his or her plans to succeed, posing the dramatic (and planning) problem of how to win their friendship back again.

Thus, a number of different knowledge representations are required to support VA behaviour, ranging from role specification, history (the VA's 'memory of events'), emotional system, and flexible goal-setting and planning abilities. Planning formalisms and techniques should thus support our implementation, though at this stage we are not committing ourselves to any specific planning approach.

However, as representations, these plans should have some specific characteristics:

- They must contain variables (i.e. they cannot be fully scripted or compiled). These variables, often the low-level operations of the plan, get instantiated according to the resources available in the virtual environment
- Their instantiation confers a new narrative meaning to the plan; the instantiated parts (terminal actions) are amenable to redescription in terms of narrative functions. This is essentially required to get a global understanding of the story when developing the system (some kind of high-level "symbolic debugging").
- More specifically, potential dispatcher functions that arise from the construction of actor's plans generate their own diagnostic procedures for the recognition of their consequences and the instantiation of the storyline

Furthermore, user intervention may be interpreted in terms of plan alteration. The user actions are altering resources for the artificial actors' plans. In some cases, the user conspicuously counters an actor's plan (or counter-plans against the actor), but this is done more by the recognition of critical actions than by a recognition of the plan as a whole (in other words this relies on background knowledge in terms of narrative functions, rather than an understanding of the system's internal state). For instance, in Figure 2 the VA has elaborated a plan that uses as a resource the letter on the set (1). If the user grabs the letter (2), it has to come with a new plan (3) using different resources (4). In addition, the level of plan alteration depends on the story generated so far.

Another important requirement is the ability to interleave planning and action execution, for instance as implemented in the MACTA planning system [17].

5. Modalities of User Intervention

We can summarise the user's modes of intervention into the story as follows: (i) the modification of the parameters of the setting; (ii) action bearing on key objects (wallet, revolver, letter...) iii) modifying the cognitive status of actors and iv) influencing VA's to take or stay away from specific actions. However, it is not only the type of intervention that is important, but also its temporal structure. Unlike the VA's who 'permanently inhabit' their Virtual World (VW), the user has a real life elsewhere and must inevitably dip in and out while still maintaining a convincing virtual existence.

Action on key objects is a straightforward physical way to influence the preconditions of key narrative actions. If we consider a situation of intense conflict between two characters, it might result in murder if a handgun is on stage and spotted by one of the actors. In a linear narrative, this would have been arranged a priori and sometimes a close-up on the alleged weapon would have signalled it to the audience. In the case of emergent narratives, only the underlying principles, i.e. the cognitive states of the characters are represented. The plans generated from these cognitive states are a result of available resources for action found in the environment.

Interfering with the set hence modifies low-level execution resources; it is up to VA's concerned to generate alternative actions. These can simply involve alternatives at the same level than the original action (i.e., picking up a paper cutter from the desktop instead of a gun). In other cases, they will involve significant differences with possibly a change in goals.

Modifying the cognitive status of actors is achieved through communication, either human-agents or agent-agent as part of the communicative actions that belong to their role plans. On the other hand, actors can similarly be influenced in real-time at key stages of the story (perceived as dispatchers by the user). This actually interferes with the plan execution. This is a more complex situation than those described above because it happens in real-time and requires that the VA is able to generate a meaningful alternative plan.

Users in the "Holodeck" can be active participants (on-stage) or spectators (being able to manipulate objects in the environments). But how can users who are off-stage still have a believable role in the ongoing action off-stage which might even be unfolding when the users are absent? If the user is in total control of the action, in the role of a director who can manipulate the script and the actors at will, the virtual agents (VA's) become no more than "puppets-on-a-string", and are not likely to be believable characters. In the real world, humans are quite sensitive to detect inconsistencies and 'odd' behaviour in fellow humans, e.g. when others are behaving 'out-of-character'. The human social world is organised around expectations and predictions of how others behave, based on previous experience with the same or similar characters. On the other hand, in the real world people's characters and roles are not static, people and their behaviour change, this is why humans have developed such elaborate skills of monitoring and predicting other's behaviour, even when the other is known for many years. On the other hand, unpredictable, spontaneous behaviour is often seen as a positive "surprise", and a sign of autonomy. As human beings we cannot directly change other's minds and 'rewrite' their beliefs and goals. Influencing (and manipulating) others requires the (at times laborious, but often enjoyable) means of communication and interaction, negotiation, discussion etc. This 'autonomy of mind' makes other human beings interesting, more interesting than the predictable behaviour of usual (fixed) characters in computer game scenarios. Unless a fixed hierarchy of command and control exists (as it is the case in human military and working organisations), the degree to which we can influence others depends on the degree and frequency of our interactions with them. Thus, the degree of social engagement can give a measure for the extent of directing others behaviour. Based on such discussions on social interaction in 'real life' we suggest the following method of user intervention in our "Holodeck" scenario when the user is "off-stage", based on the familiar "mobile phone" metaphor of communication:

When the user is off-stage, every VA in the virtual world can be reached via a communication link ('mobile phone' or 'com batch'). When the user is not on-stage he/she still has the 'social reality' for the VA's, they do not 'forget' the user, i.e. they keep a representation of the user in their memory of social agents. When off-stage the user can thought of being 'away', but occasionally calling 'home' and giving 'advice' on what decision the VA should make in its current situation. VA's answer the phone but the way the communicate with the user depends on the user's social relationship with them: a user which spends very little time on-stage in the VW and is therefore hardly known cannot expect the VA's to listen to him. In that case the VA will answer the phone but not being influenced by what the user says. On the other hand, if the user is well known to a particular VA, this VA will act upon the user's advice. Thus, it is up to the user to which degree he/she can influence the VA's when off-stage. A user who spends a lot of time in the VW and interacts with the VA's will have more influence on the plot than a user, which only occasionally stops by in the VW. The degree of influence changes dynamically, starting from very low when the "Holodeck" story is started. The user is therefore strongly motivated to spend time on-stage in the VW in order to be able to direct the story. In its simplest case the 'degree of familiarity' can be recorded by keeping track of duration and frequency of encounters between VA's and the user. The 'value' of the interaction (positive or negative) could also be considered. The role of the user as director/participant of the unfolding story in the VW is therefore dynamically influenced by the user's social immersion in the VW.

5.1. Temporal Considerations

We argue in this paper for a bottom-up interactive narrative in which the user cannot 'turn the clock back' - they remain committed to the character and actions they have carried out. The reason for this is that the ability to return to earlier saved states and make different choices, characteristic of games, is a destroyer of narrative coherence. Creating the type of VA believability which will make the user feel they are really in the virtual world also suggests the need for the concept of 'on-stage' and 'off-stage' action. This is to say that action can take place that the user does not see or directly participate in, involving only the VA's. This is necessary to provide the richness of background a convincing VA must draw on. Rather than specify this background it can also be created dynamically. Thus the narrative the user imposes on his or her experience with the virtual world is only one of many possible such "storifications". A continuously running virtual world, in which events occur that the user is not involved in poses some interesting issues.

Firstly, the narrative is likely to be episodic in character. There are several reasons for this: one is that the user is only present in the narrative (or even as an observer) some of the time. Another is that if activity is to take place in different physical locations, a change of scene must be produced without the kind of non-narrative effort involved in getting onto a virtual bus. This is a specific case of the general consideration that not all activity is intrinsically interesting, in which case it should happen off-stage.

A user who is not permanently present, must have an "in-character" method of joining and leaving the world (an analogy with the group-controlled skipping rope of childhood suggests itself - the user must jump in and out neatly without getting tangled in the rope!). This has implications for the role that the user is allocated in the first place - if they play character 'K', there should always be an in-character answer to any VA who asks 'where is K?', and, more strongly, an in-character answer from 'K' if asked directly 'where were you yesterday?'.

Next, 'K' must only be able to leave the world in a justifiable fashion. For example, not in mid-sentence; not just as some bad action meets its just deserts from a VA. It may be necessary to incorporate some exit mechanism such as a special door leading 'off-stage' for a

VA and out of the world altogether for the user. The same applies to entering the world - K should not suddenly appear in the middle of a group of VA's 'as if by magic' unless the scenario supports magic appearances and the VA's can also make use of it. The guiding point here is that the world must possess a guiding logic, which applies as much to the user as to the VA's.

The introduction of 'off-stage' action requires a mechanism for communicating significant events back to the user. Both in-world and global mechanisms can be considered: the former could be embodied in a VA who gossips about what has happened or in a newspaper or virtual radio/TV (or a virtual website?). The latter could take the form of a voice-over summary of the type used at the start of TV episodes (or in written adverts for TV soaps) which would update the user if they had been absent from the world for some time.

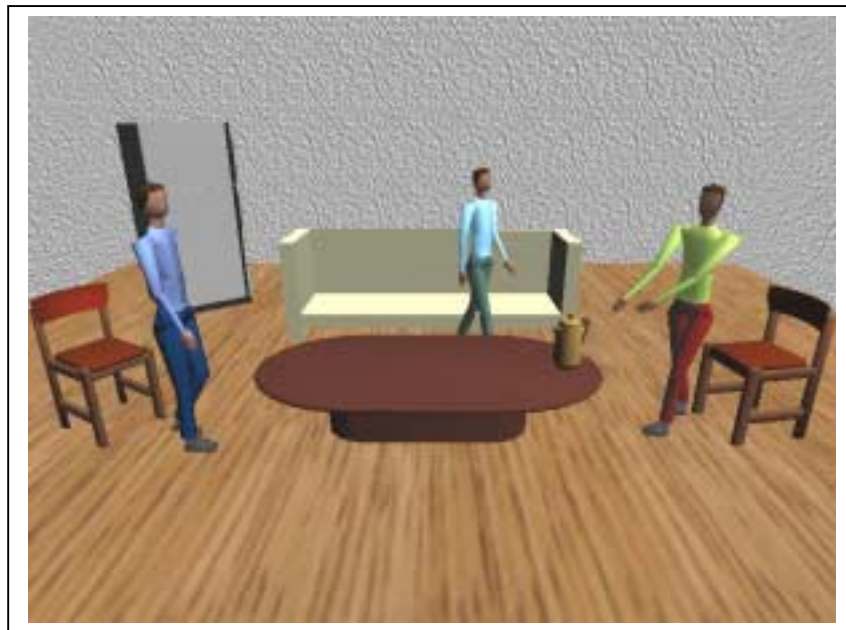


Figure 3. A Sample Set Developed with the *Maverik* System.

This would allow K to reappear in an appropriate place in relation to his or her view of the continuing narrative.

In observer mode, the user might be allowed to follow a VA from scene to scene. Here the issue is of guiding the user towards the interesting pieces of off-stage activity (by definition, potentially on-stage if the stage moves) and away from characters sitting eating their dinner on their own and reading a newspaper. Forcing the user to adopt a specific observation position is all too much like current non-interactive media, while letting the user wander around at will risks narrative coherence - some intermediate position in which appropriate guidance and choices are offered is probably the way forward.

6. User Interaction Techniques

The interaction techniques should support natural modalities and temporal structure and implement the modes of intervention we have described in previous sections, i.e. on-stage and off-stage.

The physical interaction with the virtual set objects can be supported through various motion tracking devices. The user can be equipped with sensors, though this makes less natural his transition from on-stage participant to spectator. However, physical interaction is more natural in an immersive environment such as a CAVE.

Clearly, interacting with actors is most natural through speech and spoken dialogue. In narrative terms, this would provide a unified framework for user intervention, as speech acts would match physical intervention. However, speech acts described in conversational analysis are much too generic and we shall instead devise specific speech acts. These would be ideally mapped to the structure of narrative functions. This is why the project will use speech understanding and dialogue technologies developed previously in [16]. These can be used both in immersive and semi-immersive environments (such as large-screen projection systems). However, there are limitations both in the size of vocabulary and in syntactic coverage.

One possibility is to use multi-keyword spotting to give the user a largely unconstrained mode of expression. This would have the advantage of accepting spontaneous speech that include interjections, hesitations, etc.

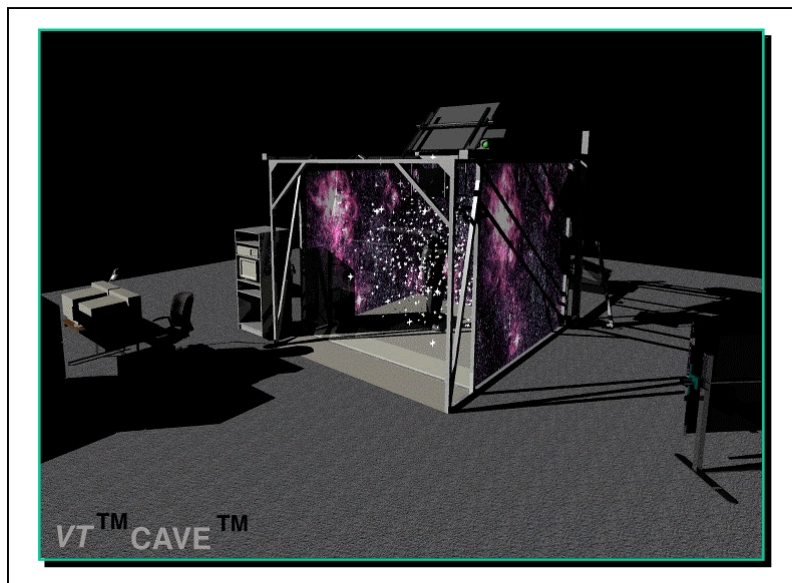


Figure 4. A CAVE™ Installation.

The rationale for the use of multi-keyword spotting in a storytelling context is that it has the ability to spot topics (whether these are story objects, characters or events) and reconstruct simple functional relations between them. This approach however limits the complexity of direct orders that could be passed to actors. The spoken dialogue technology we developed previously can accommodate the multi-keyword spotting paradigm and should be re-used to implement spoken interaction.

7. Conclusions

We have introduced some fundamental problems behind the implementation of a “Holodeck”. The system will be based on a CAVE™ immersive display (Figure 4), in which a single human user will interact with artificial actors and with objects of the virtual environment. We’ll be experimenting with several VR environments, but are currently using the Maverik system [19] (Figure 3). Obviously, as in many IVE experiments, we are concentrating on behaviour rather than photorealism.

This project would not be at all possible without the re-use of previous technology for spoken interaction, human-computer dialogue [18], planning [17]. Even though these background technologies have been fully implemented, the “Holodeck” itself is still in his

very early days. Our first experiments should be dedicated to the integration of planning techniques in a simple world featuring a few artificial actors in a simple, “sitcom-like” scenario.

References

- [1] E.J. Altman, and R. Nakatsu. Interactive Movies: Techniques, Technology and Content. Course Notes n. 16, ACM SIGGRAPH'97, 1997.
- [2] J.H. Murray, *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. MIT Press, Cambridge, 1997.
- [3] R. S. Wyer, Ed. *Knowledge and Memory: the Real Story*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1995.
- [4] J. Bruner. The Narrative Construction of Reality, *Critical Inquiry* 18(1), 1-21, 1991.
- [5] K. Nelson. The psychological and social origins of autobiographical memory, *Psychological Science* 4(1), 7-14, 1993.
- [6] K. Dautenhahn. The Lemur's Tale - Story-Telling in Primates and Other Socially Intelligent Agents. *Proc. Narrative Intelligence, AAAI FS 1999*, Cape Cod, AAAI Technical Report FS-99-01. 59-66, 1999.
- [7] D.C. Dennett. The Origin of Selves. *Cogito* 3, 163-173, 1989.
- [8] R. Aylett. Narrative in Virtual Environments - Towards Emergent Narrative. *Proc. Narrative Intelligence, AAAI FS 1999*, Cape Cod, AAAI Technical Report FS-99-01. 83-86, 1999.
- [9] K. Dautenhahn. Story-Telling in Virtual Environments. *Proc. Workshop on Intelligent Virtual Environments, ECAI-1998*, Brighton, 1998.
- [10] M. Umaschi Bers, J. Cassell. Children as Designers of Interactive Storytellers “Let me tell you a story about myself...”. In K. Dautenhahn (Ed.) *Human Cognition and Social Agent Technology*, John Benjamins Publishing Company, 61-83, 2000.
- [11] P. Sengers. Do the thing right: an architecture for action-expression. *Proc. of the Second International Conference on Autonomous Agents*, Minneapolis/St. Paul, USA, May 9-13, 24-31, 1998.
- [12] B. Webber, N. Badler, B. Di Eugenio, C. Geib, L. Levison, L., and M. Moore. *Instructions, Intentions and Expectations*. Technical Report, IRCS-94-01, Institute for Research In Cognitive Science, University of Pennsylvania, 1994.
- [13] R. Raskin. Five Parameters for Story Design in the Short Fiction Film. *P.O.V.*, n. 5, 1998.
- [14] R. Barthes, *S/Z*, Paris: Seuil, 1966.
- [15] R. Barthes, Introduction a l'analyse structurale des récits, *Communications*, n. 8, 1968.
- [16] Cavazza, M. and Palmer, I.. High-Level Interpretation in Virtual Environments. *Applied Artificial Intelligence*, 14:1, pp. 125-144, 1999.
- [17] Aylett, R.S.; Coddington, A.M.; Barnes, D.P.; Ghanea-Hercock. Supervising multiple cooperating mobile robots. *Proceedings, 1st International Conference on Autonomous Agents*, Marina del Rey, 1997
- [18] M. Cavazza. VIP: A Conversational Character for Interactive Television. *Third Workshop on Human-Computer Conversation*, Bellagio, Italy, 2000.
- [19] T.L.J. Howard, R.J. Hubbold and A.D. Murta, MAVERIK: A Virtual Reality System for Research and Teaching, *GVE'99*, Coimbra, Portugal. 1999