LNCS 2197

Olivier Balet
Gérard Subsol
Patrice Torguet (Eds.)

Virtual Storytelling

Using Virtual Reality Technologies for Storytelling

International Conference ICVS 2001
Avignon, France, September 2001
Proceedings

Springer

Lecture Notes in Computer Science 2197
Edited by G. Goos, J. Hartmanis and J. van Leeuwen

Springer
Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Tokyo

Olivier Balet   Gérard Subsol
Patrice Torguet (Eds.)

Virtual Storytelling

Using Virtual Reality Technologies for Storytelling

International Conference ICVS 2001
Avignon, France, September 27-28, 2001
Proceedings

Springer

Volume Editors

Olivier Balet
CS SI, Virtual Reality Department
ZAC de la Grande Plaine, rue Brindejonc des Moulinais
31029 Toulouse, France
E-mail: Olivier.Balet@c-s.fr

Gérard Subsol
Université d'Avignon, Laboratoire d'Informatique
339, chemin des Meinajariès, 84911 Avignon Cedex, France
E-mail: Gerard.Subsol@lia.univ-avignon.fr

Patrice Torguet
Université Paul Sabatier, IRIT
118, route de Narbonne, 31062 Toulouse Cedex 4, France
E-mail: torguet@irit.fr

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Virtual storytelling : using virtual reality technologies for storytelling ; proceedings / International Conference ICVS 2001, Avignon, France, September 27 - 28, 2001. Olivier Balet ... (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; London ; Milan ; Paris ; Tokyo : Springer, 2001
(Lecture notes in computer science ; Vol. 2197)
ISBN 3-540-42611-6

CR Subject Classification (1998): I.3, I.2, C.3, H.4, H.5, I.7.2, I.4

ISSN 0302-9743
ISBN 3-540-42611-6 Springer-Verlag Berlin Heidelberg New York

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP Berlin, Stefan Sossna
Printed on acid-free paper SPIN 10840703 06/3142 5 4 3 2 1 0

Preface

The story is the richest heritage of human civilizations. One can imagine the first stories being told, several thousand centuries ago, by wise old men huddled around campfires. Since this time, the narrative process has been considerably developed and enriched: sounds and music have been added to complement the speech, while scenery and theatrical sets have been created to enhance the story environment. Actors, dancers, and technicians have replaced the lone storyteller. The story is no longer the sole preserve of oral narrative but can be realized in book, theatrical, dance, or movie form. Even the audience can extend up to several million individuals.

And yet in its many forms the story lies at the heart of one of the world's most important industries.

The advent of the digital era has enhanced and accelerated this evolution: image synthesis, digital special effects, new Human-Computer interfaces, and the Internet allow one not only to realize more sophisticated narrative forms but also to create new concepts such as video gaming and virtual environments. The art of storytelling is becoming evermore complex. Virtual reality offers new tools to capture, and to interactively modify the imaginary environment, in ever more intuitive ways, coupled with a maximum sensory feedback. In fact, virtual reality technologies offer enhanced and exciting production possibilities for the creation and non-linear manipulation in real time, of almost any story form. This has lead to the new concept of **Virtual Storytelling**.

The first International Conference on Virtual Storytelling gathers researchers from the scientific, artistic, and industrial communities to demonstrate new methods and techniques, show the latest results, and to exchange concepts and ideas for the use of Virtual Reality technologies for creating, scripting, populating, rendering, and interacting with stories, whatever their form, be it theatre, movie, cartoon, advertisement, puppet show, multimedia work, video-games...

We hope that ICVS 2001 will be of great interest to all the participants and that it will be the first conference in a long series of international conferences on this fascinating topic.

September 2001

Olivier Balet
Gérard Subsol
Patrice Torguet

Acknowledgements

The creation of this international conference on Virtual Storytelling is a joint initiative of the Virtual Reality Department of the *Communication et Systèmes* Group and the French Working Group on Virtual Reality (GT-RV).

However, this conference would not have existed without the official support of the European Commission and the active contribution of sponsors, organizations, and individuals.

Therefore, the conference organizers would like to thank the European Commission's IST Program for contributing to the conference funding as well as both the *Immersion SA* and *RealViz* companies for offering wonderful prizes to the best papers.

We would also like to thank the members of the Scientific and Application Boards for supporting the conference from the very beginning, helping us to indentify the most relevant topics, and proposing names of experts to sit on the Program Committee.

The members of the Program Committee deserve special acknowledgment for their superb job in reviewing all the papers with such remarkable care and moreover... by the deadline!

Finally, we wish to thank the people from *Sophie et Associés* and the Computer Science laboratories of the Universities of Avignon and Toulouse for their help during the preparation of this conference.

Organization

Chair

Olivier Balet	Communication & Systèmes Group, France
Gérard Subsol	LIA, University of Avignon, France
Patrice Torguet	IRIT, University of Toulouse, France

Scientific Board

Bruno Arnaldi	SIAMES Project, IRISA, France
Eric Badiqué	European Commission, Brussels, Belgium
Norman I. Badler	University of Pennsylvania, USA
Andrew Berend	Anthropics Technology Ltd., UK
Bruce Blumberg	Media Lab, MIT, USA
Michael F. Cohen	Microsoft Research, Redmond, USA
Enrico Gobbetti	CRS4, Italy
Martin Göbel	GMD, St Augustin, Germany
Alain Grumbach	ENST Paris, France
Barbara Hayes-Roth	Stanford University/Extempo Systems, USA
Ryohei Nakatsu	ATR, Kyoto, Japan
Philippe Quéau	UNESCO, Paris, France
Paul Siebert	University of Glasgow, UK
Kristian Simsarian	IDEO, San Francisco, USA
Frank Van Reeth	Limburgs Universitair Centrum, Belgium

Application Board

Jean-Francis Balaguer	Largeur, Switzerland
Maurice Benayoun	Z-A Productions, Paris, France
Franz Fischnaller	F.A.B.R.I.CATORS, Milano, Italy
Andrew Glassner	Writer Director, USA
Tran Anh Hung	Director, Vietnam
Paul Kafno	HD Thames, UK
Mathieu Kassovitz	Director, Actor, France
Sally Jane Norman	Ecole Supérieure de l'Image, Angoulème, France
David Parfitt	Producer, Actor, UK
Ulrike Spierling	ZGDV, Darmstadt, Germany
Vittorio Storaro	Cinematographer, Italy

Program Committee

Edward Altman	Kent Ridge Digital Labs, Singapore
Cynthia Breazeal	Sociable Machines Project, MIT, USA
Marc Cavazza	University of Teesside, UK
Jean-Luc Dugelay	EURECOM Sophia Antipolis, France
John Funge	iKuni Inc., Palo Alto, USA
Petra Gemeinboeck	University of Illinois at Chicago, USA
Jonathan Gratch	University of Southern California, USA
Jean-Pierre Jessel	IRIT, France
Vali Lalioti	University of Pretoria, Republic of South Africa
Wim Lamotte	Limburgs Universitair Centrum, Belgium
Keith Miller	University of Illinois at Chicago, USA
Sina Mostafawy	rmh, Köln, Germany
Jean-Christophe Nebel	University of Glasgow, UK
Tim Regan	Microsoft Research, Redmond, USA
Agnès Saulnier	Institut National de l'Audiovisuel, France
Nikitas Sgouros	University of Piraeus, Greece
Danaë Stanton	University of Nottingham, UK
Marie-Luce Viaud	Institut National de l'Audiovisuel, France
R. Michael Young	North Carolina State University, USA

Table of Contents

New Techniques

Authoring Tools

A New Form of Narration

Virtual Characters

Applications

Under Construction in Europe: Virtual and Mixed Reality for a Rich Media Experience

Eric Badiqué

European Commission, IST Programme
Eric.Badique@cec.eu.int
www.cordis.lu/ist/ka4/vision/

Abstract. As part of the 5th Framework Programme, the European Commission supports research and development activities in the field of Virtual and Mixed Reality (VR&MR), in particular within the Information Society Technologies (IST) specific Programme. This paper gives an overview of VR&MR and related projects already launched. Finally, some very preliminary orientations for future research are given.

1 The European IST Programme

The Fifth Framework Programme defines the Community activities in the field of research, technological development and demonstration for the period 1998-2002. It focuses on a limited number of objectives and areas combining technological, industrial, economic, social and cultural aspects. The Fifth Framework Programme consists of seven Specific Programmes. An overview of the 5th Framework Programme is available under [1].

The Information Society Technologies (IST) Programme is part of the FP5. It is based on the premise that the Information Society is about to move into an era where technology will be all around us but almost invisible and where networked devices embedded in commonplace appliances enable people to have easier interactions with services. The Programme focuses its activities on the realisation of a "vision" that is user-centred: "*Our surrounding is the interface*" to a universe of integrated services. While directly targeting the improvement of quality of life and work, the vision is expected to catalyse a wealth of business opportunities arising from the aggregation of added value services and products.

This vision promotes both ubiquity and user-friendliness of IST and focuses on the combination of the two concepts into "ambient intelligence" environments. Realisation of the vision presents many technical challenges, including issues related to VR, MR, advanced interfaces and Digital Storytelling. These issues are, in particular, developed in the Action Line 'Mixed Realities & New Imaging Frontiers'.
A full description of the IST Programme is available under [2] and [3].

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 3–9, 2001.

2 The Action Line 'Mixed Realities & New Imaging Frontiers'

2.1 Objectives and Scope

The objective of this Action Line is to bridge the gap between real and virtual worlds for innovative applications. The focus is on the Reality-Virtuality continuum:

- Augmenting virtuality and bringing virtual worlds to life by enhancing realism and level of detail, introducing intelligence, making them persistent and reactive environments;
- Augmenting reality and fusing real and virtual universes by enhancing real environments for applications ranging from wearable computing for navigation and industrial processes to programme production and interactive entertainment; and
- Discovering new sensory frontiers by addressing high definition, 3D, full space imaging, multisensory cues and very advanced display systems to create fully immersive environments distributed over heterogeneous networks and platforms in which users will be able to enjoy rich, multisensory experiences for virtual- or tele-presence.

Costs, real-time, human factors, control, protection and ethical issues are also important aspects when preparing a proposal and should be considered.

Mixed Reality (MR) makes the best out of two worlds, effectively marrying the flexibility of computer graphics with the realism of real-life pictures. For this, computer vision, computer graphics and advanced audio-visual representation and coding techniques need to be integrated. There is little doubt that within the not too distant future, MR will lead to new visual interfaces, which are needed to move beyond the desktop paradigm. MR is not limited to visualisation and should also be seen in a wider context, opening the way to the integration of mechanics, robotics, toys and appliances with visualisation and IT equipment. Future MR applications will seamlessly integrate real world objects one can touch and feel with software and audio-visual representations. More details about this action line are available from [4].

2.2 Running Projects

17 projects have so far been accepted for funding under this action line. The following table gives an overview. More information is available from [5].

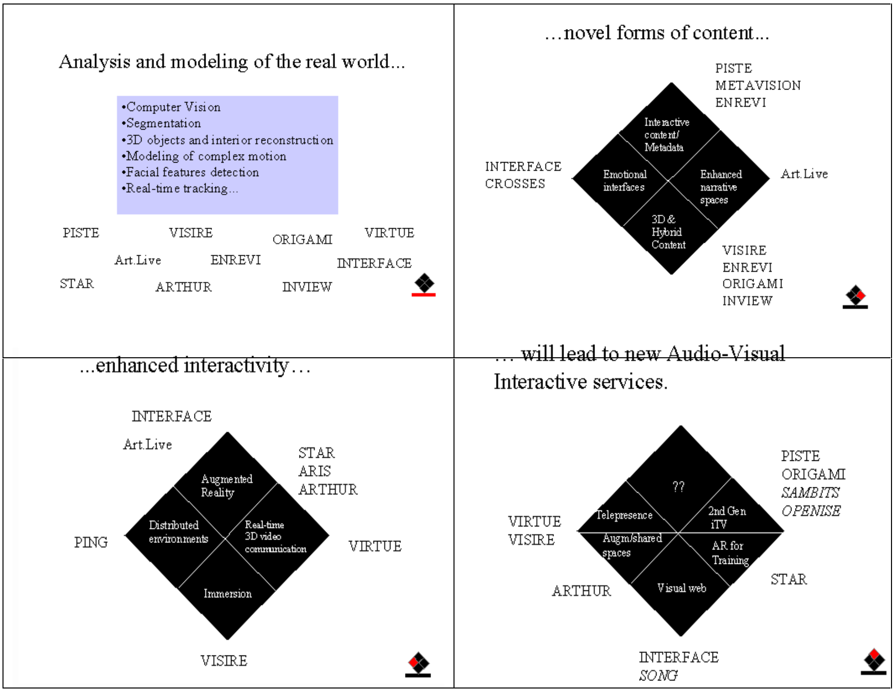
There is also a significant number of projects related to VR and Digital Storytelling in other parts of the Programme with an emphasis on long-term research or applications and content. See for example the Action Lines 'Publishing digital content', 'Next generation digital collections' or 'x-Content futures'. See [6], [7] and [8] for references.

Table 1. Funded projects

IST-28707	ARIS	Augmented Reality image synthesis through illumination reconstruction and its integration in interactive and shared mobile AR systems for E-(motion)-commerce
IST-28559	ARTHUR	Augmented round table for architecture and urban planning
IST-10942	Art.Live	Architecture and authoring tools for prototype for Living Images and new Video Experiments
IST-10510	CROSSES	CROwd Simulation System for Emergency Situations
IST-11185	ENREVI	ENhanced REality for the Video
IST-10036	INTERFA-CE	Multimodal Analysis/Synthesis System for Human Interaction to Virtual and Augmented Environments
IST-28459	INVIEW	Interactive and immersive video from multiple view images
IST-20859	META VISION	Universal Electronic Production system
IST-28436	ORIGAMI	A new paradigm for mixing of real and virtual
IST-11488	PING	Platform for Interactive Networked Games
IST-11172	PISTE	Personalised, Immersive Sports TV Experience
IST-11683	SAFIRA	Supporting Affective Interactions for Real-time Applications
IST-28764	STAR	Services and training through Augmented Reality
CRAFT 99-56418	SYMUSYS	Innovative High-performance Motion Simulation System For Entertainment, Research And Training Applications
IST-10044	VIRTUE	VIRtual Team User Environment
IST-10756	VISIRE	Virtual Image-Processing System for Intelligent Reconstruction of 3D Environments
IST-20783	VRSUR	Virtual Reality Surgery Training System

2.3 Coverage

The projects in this cluster are positioned at the cross-road of content and interfaces. Some provide generic frameworks for leveraging computer vision, AV representation, computer graphics, agents and mixed reality technology. The diversification of the projects' target implementation platforms ranging from embedded "smart" cameras to second-generation set-top-boxes is testimony that we are moving into the "post-PC" era. These projects also strive to keep a good balance between theoretical research and dealing with practical constraints. Collectively, the projects contribute to the development of novel audio-visual interactive services as illustrated below:



2.4 Strategic Impact

The exploitation of mixed reality and visualisation technologies will arguably be the next step of development of the information infrastructure as we know it today (WWW, dTV and mobile). These technologies will give us new interfaces for the "nintendo&playstation generation", the generation used to interact with highly visual and interactive systems, such as the games they grew up with. They will offer new images and new forms of content, new ways of learning, new ways of caring, new ways of working, new ways of making business, new opportunities and new markets.

Beyond the hype and the stunning images MR is able to produce, there are already many very practical and high potential applications foreseeable in fields such as maintenance, medical visualisation, guidance and information, entertainment, e-learning, traffic, architecture or office communication. All these applications are currently being explored in collaborative projects.

But the EU agenda goes beyond technology and markets to include societal and socio-economical issues important for the development of the "information society". Since it enables more natural and intuitive interfaces, MR could have a large impact on usability and access to services. By lowering the access threshold, it could consequently contribute to bridging the "digital divide" and favour a wider access to information, every-time and every-where. Through its contribution to the visualisation of large amounts of information, edutainment and learning by doing, it could also revolutionise education and training.

There are still a number of key technical issues to resolve before reliable and acceptable MR systems can be deployed. The integration of mobility, real-virtual visualisation aspects and communication is one of the keys. The development of light and non-intrusive sensors and displays is another. Analysis, understanding and tracking are paramount. More work is needed on human factors aspects.

Once these problems are overcome, a technology such as MR is likely to lead to the creation of completely new markets. In some manufacturing industries, it has the potential to be the next wave of innovation following CAD and robotics. In this context, it may lead to the "revaluation" of manual work. In the media area, MR will be very appealing to service providers in need of differentiation and to operators in need of bandwidth-hungry applications. MR is also likely to be an integral part of the next generations of interactive TV, a mixture of broadcasting and Internet services.

By making it easier to find, use and share information, and by offering new, attractive service options, these technologies will provide fuel for the digital economy and help it achieve a sustained growth. The development potential is enormous.

3 Perspectives – From Information to Experience?

The European Commission is currently proposing orientations for the next Framework Programme covering the period 2002-2006. At the time of writing, the work programme is still under discussion and no conclusions have been reached concerning the priorities and the implementation instruments.

However, it is likely that technologies related to VR, MR, interfaces and digital storytelling will continue to be supported since they directly underpin the development of advanced interfaces essential to a “user-friendly” information society.

These technologies are likely to appear in R&D work on intelligent surfaces and interfaces, aimed at developing more effective ways of accessing ubiquitous information and natural interaction modes. They will contribute to interfaces that are intuitive, adaptive and multi-sensorial, and that will hide the complexity of technology by supporting a seamless human interaction with devices, virtual and physical objects, in a variety of environments.

They will also be essential to research in knowledge technologies and digital content aiming at automated solutions for creating, managing and interacting with complex knowledge spaces.

Finally, they will play an important role for communication technologies in which research will address the enabling building blocks for personalised access to networked audio-visual systems and applications and will target appliances able to process, encode, store, sense and display hybrid rich-media signals and objects [9].

Without attempting to be exhaustive, the key core technological challenges are likely to include [10]:

- Computer vision for capturing and reconstructing static and dynamic scenes and for generating virtual copies of real environments and objects to enable real-time tele-immersion and augmented reality applications. It will also concern vision sensors, robustness to noisy (light, alignment, motion) environments, capturing and understanding of qualitative visual information, such as gestures, facial expressions, behaviors and contextual information.
- Audio-visual data manipulation and storage, including indexing and manipulation of information, cross-modal descriptions, representation and coding, computer graphics, Mixed Reality technologies, digital storytelling, metadata production tools, as well as new imaging and sensory frontiers for immersive media.
- Language and speech in order to understand the user, multilinguality, taking into account the recipient's preferences or linguistic skills and the emotional context as well as robustness to noisy environments.
- Haptic and other sensory interfaces including touching and feeling virtual objects, other senses such as odor and taste as well as localized and spatialised audio.
- Context awareness and self-awareness for context-sensitive adaptation including self-awareness, emotive contexts and understanding of other people the user interacts with.
- Adaptivity, personalization and control including learning mechanisms and work on human behavior.
- Human/human interaction and human/machine interaction for the understanding of various aspects of human/human interaction and of the relation between people and machines.
- Displays able to render large amounts of heterogeneous information.
- Standards effective at many different technological levels (multimodal data coding, middleware, synchronization, wired and wireless connectivity, data storage and manipulation, etc).

To guarantee success, it will be essential to ensure a successful interaction between core technologies by, for example, promoting the development of synergy between natural image processing, computer vision and computer graphics or encouraging a deeper integration of vision and speech work. Work on human/human interaction and human/machine interaction should also be combined with work on language, vision and intelligent information presentation.

The concept of ***experience*** has been proposed as a powerful driver for integration of various technologies. As opposed to 'conventional' media for communication, retrieval or "consumption", enjoying an experience is strongly related to being able to re-establish context. It is also about enhanced interactivity and advanced user interfaces involving multi-sensor(y) data collection including continuous, real-time acquisition and tracking of events. A real experience requires powerful databases and knowledge management systems as well as effective ways to generate and distribute metadata for personalisation. Finally, it needs high-quality visualisation technologies including 3D, VR, Mixed Reality and immersive displays.

Experiences are about content that is highly personalised, interactive, contextual and event-centric. It is based on the fact that in our daily life, we use our five senses to acquire continuous data about the world around us, and we assimilate this data with information and knowledge from the past to understand and respond to our environment. It is about organising information in a way that people experience, by time and space and around concepts, not keywords [11]. Experiences reach beyond information and could be seen as the ultimate media.

There are already some examples of first experiences becoming available: Portals for live sport events including tracking and notification, based on a complete integration with databases for past and future events go beyond simple sport reporting; The 24 hours tracking of boats during a race around the world provides a first racing experience ; The tracking of Formula 1 cars and their inclusion in a game in real-time, effectively provides more than the sum of broadcasting and gaming.

This is only the beginning. With the help of a European Programme such as IST, a few well targeted future projects could integrate and further develop the best tools available today and invent the next generation of digital experiences, for all of us to enjoy.

Acknowledgements. The author wishes to thank Ramesh Jain of UCSD's Visual Computing Laboratory for an inspiring talk, as well as all the experts who have contributed to the recent FP6 consultation meetings held in Brussels in Spring 2001.

Disclaimer. The opinions expressed in this paper are those of the author and do not necessarily reflect the views of the European Commission.

References

1. www.cordis.lu/fp5/src/over.htm
2. www.cordis.lu/ist/overview.htm
3. www.cordis.lu/ist/workprogramme.htm
4. www.cordis.lu/ist/ka4/vision/home.html
5. www.cordis.lu/ist/ka4/vision/projects.htm
6. www.cordis.lu/ist/fethome.htm
7. www.elpub.org/
8. www.cordis.lu/ist/ka3/digicult/home.html
9. IST Consultation Meeting on Communication Infrastructures, Brussels, 2-3 May 2001 (internal document)
10. Multisensorial, Multilingual Interfaces and Virtual Environments, Report of the IST Program Consultation Meeting 10, Brussels 26-27 April 2001 (internal document)
11. http://www.praja.com/

Generation of True 3D Films

Jean-Christophe Nebel

3D-MATIC Research Laboratory, Department of Computing Science,
University of Glasgow, Glasgow, Scotland, UK
jc@dcs.gla.ac.uk

Abstract. We define a true 3D film as a film that can be viewed from any point in space. In order to generate true 3D films, the 3D-MATIC Research Laboratory of the University of Glasgow has been developing a capture 3D studio based on photogrammetry technology. The idea is simply to generate 25 photo-realistic 3D models of a scene per second of film. After the presentation of the state of the art in the domain, the core technology of our dynamic 3D scanner is detailed. Finally first results, based on a 12 camera system, are shown and the potential applications of this new technology for virtual storytelling are investigated.

1 What Is a True 3D Film?

From the 30s onwards cinema has tried to create the illusion of the 3D image by artificially reproducing binocular vision. The cinema aims to film separately the same object from two angles corresponding to our binocular vision and to project them onto the same screen. It only remains for each eye to select the image, it is meant to receive, in order to recreate a 3D illusion. Special glasses are used for this purpose. This technology has been quite successful considering that nowadays such 3D films are offered every day in cinemas such as IMAX^{TM}.

Since these films are filmed from a specific viewpoint, spectators can only see 3D images from that viewpoint. For some people, it may be a bit frustrating, they would like to see these 3D images from other view points: they would like to see true 3D film. We define a "true" 3D film as a film that can be viewed from any point in space. Ideally spectators should have the ability to choose interactively the position of the camera while watching these films; they should be able to fly over scenes of a film, as it is now possible to navigate through 3D virtual reality environment using VRML browsers. In order to generate these true 3D films, the 3D-MATIC Research Laboratory of the University of Glasgow has been developing a capture 3D studio based on photogrammetry technology.

In this paper, we present the prototype of the 3D studio we are currently developing. First we relate our research to previous work in the fields of 3D capture, modelling and animation of virtual humans. Then we describe the technology we developed and finally we show some experimental results and offer some applications for virtual storytelling.

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 10–19, 2001.

2 Related Work

In spite of being a very active research topic, convincing animations of realistic virtual humans have been demonstrated only in very few short films such as "Tightrope" by Digital Domain^{TM} and "L'Opus Lounge" with "Eve Solal" by Attitude Studio^{TM}. Two main challenges have to be addressed in order to generate such films: the creation of photo-realistic 3D models of real humans and the realistic animation of these models.

2.1 Creation of Photo-Realistic 3D Models of Real Humans

The first challenge is the creation of photo-realistic 3D models of real humans. On one hand skilled designers are able to make human models using software such as 3D Studio Max^{TM}. However since few months are necessary for the creation of a convincing model, they generally represent average humans, specific film stars or athletes. On the other hand specific human models can be generated using automatic or semi-automatic techniques. There are mainly two main methods: the deformation of generic 3D models and the generation of 3D models using scanners. In the first case, pictures are mapped on a generic 3D model of an average character, which has been scaled and deformed in order to match the pictures. Blanz et al. generate faces from a single picture using a morphable face model [1], which was built using statistics acquired from a large dataset of 3D scans. Hilton et al. map pictures of the full body of a person, taken from 4 orthogonal views, on a generic 3D human model representing both shape and kinematic joint structure [6]. These techniques produce very realistic 3D models. However since they are based on only few pictures and a generic model, the similarity between the human model and the generated model depends on the viewpoint. The other way of generating automatically realistic humans is by using scanners. Several techniques can be used to scan a human body: laser beams [17] and Cyberware^{TM}, structured light technique [21] or photogrammetry [15] and [18]. Their accuracy (about 1mm) is usually sufficient for getting very realistic 3D models. Moreover colour pictures are mapped on these models what ensures photo realistic appearance.

2.2 Realistic Animation of Human Models

Once these photo-realistic 3D models are available, the second challenge needs to be addressed: their animation. There are many software allowing the animation of human like figures using key frames such as Maya^{TM} and Poser4^{TM} and a lot of work has been done in order to ease the way of generating poses using techniques such as emotional posturing [3], [1] and genetic algorithms [4]. However, it is still a very long task and requires highly skilled animators to generate realistic motion. Therefore research has focused more recently on high level techniques such as adapting reference movements obtained either by keyframing or motion capture. The higher level of control provided reduces the animator's load of direct specifications for the desired movement. Many approaches have been followed such as the interpolation between keyframes of reference motions [5] and [2], the generation of collision free motions [12] and the derivation of a motion from a reference motion by adding emotions or behaviours to keyframes [20] and [3]. In conclusion, a lot of work has been done in order to speed up the process of generating realistic animations, but ultimately an animator is still needed to set the fine tunings.

2.3 3D Capture of Human Motion and Shape

For the time being it seems that the only practical way of generating quickly convincing 3D animations of human beings is to use real people as much as possible: the actors should be scanned and their motions should be captured. Therefore what is needed is a 3D scanner which would be able to scan a full moving body in a very short capture time, in order to freeze the motion, and would be able to scan this body, ideally, at a cinema or TV frame rate. Very few of the scanners presented previously have a short enough capture time. The commercial scanners, based on laser beams and structured light, have a capture time of about 10 seconds, whereas the ones using photogrammetry only need few milliseconds. Obviously only the later type of scanners has the potential of capturing moving subjects. People of the research team of the British company TCTi^{TM} [16] work on the generation of true 3D movies using photogrammetry based scanning technology, however no result has been published yet.

The Robotics Institute of Carnegie Mellon University has also an interest of capturing and analysing 3D human motion. For that purpose they built a "3D room" which is a facility for 4D digitisation: a large number of synchronised video cameras [8] are mounted on the walls and ceiling of the room. Since their main interest is the analysis of human motion [19], they have not developed any advanced method in order to generate accurate 3D models. However using silhouette carving, they managed to generate sequences of crude 3D models which have allowed them to create amazing true 3D films (see [14]).

It is also worth mentioning the work by Monks [11]. They designed a colour encoded structured light range-finder capable of measuring the shape of time-varying or moving surfaces. Their main application was about measuring the shape of the human mouth during continuous speech sampled at 50Hz. Since their system is based on the continuous projection of a colour encoded structured light, their technique has some limitations compared to ours. Their structure light can only be projected from a single direction; therefore they cannot get a full coverage of 3D objects. Moreover the capture of the texture of 3D objects is not possible.

3 Principle

Since realistically believable true 3D films cannot be generated easily using animation techniques, we offer a totally new method: the capture of 3D data using scanning techniques at a frame rate of 25 scans per second. The idea is simply to generate 25 3D models of the scene per second of film. Since the capture time of the scanner has to be very fast, we use a scanner based on photogrammetry which has a capture time of few milliseconds [15]. Therefore that gives us the ability to capture subjects at a frame rate of 25 scans per second. The prototype of the 3D studio we are currently developing allows the 3D capture of a scene fitting a 2 metres side cube, typically we can capture the motion of a single actor. The configuration of this scanner is the following: the scene will be imaged by a total of 24 TV cameras arranged in threes. We term a group of three cameras a pod. Eight pods, arranged at the corners of a parallelepiped will image the active volume (see Fig. 1). For the time being only 12 cameras have been installed. A more detailed presentation of our dynamic 3D capture system is given in [13]

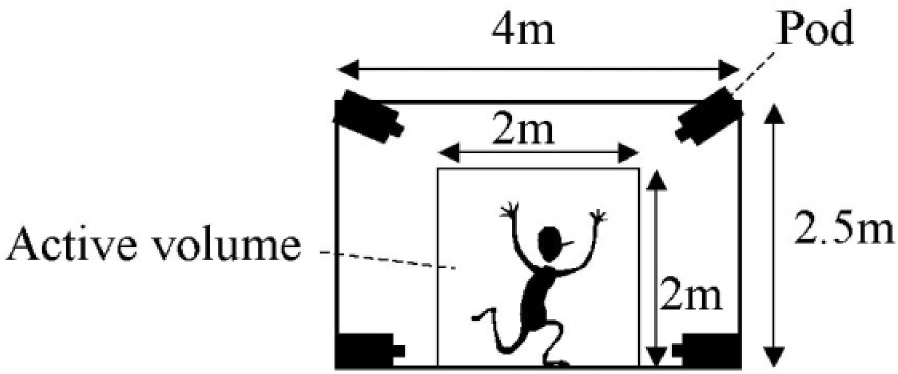


Fig. 1. Configuration of the 3D studio

4 Dynamic 3D Data Capture

The process of 3D capture relies upon flash stereo photogrametry. Each pod has one colour and two black and white cameras. Associated with each pod are two strobe lamps, one of which is a flood strobe, the other is fitted within a modified overhead projector which illuminates the scene with a random dot pattern. At the rate of 25Hz successively, the colour cameras capture the scene illuminated with uniform white light, and then the mono cameras capture the scene illuminated with the texture. The total capture time is under $150\mu s$. The monochrome images are used for stereo range finding and the colour images are used to capture the surface appearance of the subject.

In order to build 3D models from the data captured by the scanner previously described, the cameras have to be calibrated, e.g. the detailed geometric configuration of all the cameras has to be known. Then once the capture has been done, the stereo matching process is applied to each stereo-pair images. The algorithm we use is based on multi-resolution image correlation [22].

The algorithm takes as input a pair of monochrome images and outputs a pair of images specifying the horizontal and the vertical displacements of each pixel of the left image compared to the matched point in the right image (see Fig. 2). The matcher is implemented using a difference of gaussian image pyramid: the top layer of the pyramid is 16 by 12 pixels in size for a base of 640 by 480. Starting from the top of the pyramid, the matching between the 2 pictures is computed. Then using the displacements, the right image of the next layer of the pyramid is warped in order to fit the left image. Thus if the estimated disparities from matching at the previous layer were correct, the two images would now be identical, occlusions permitting. To the extent that the estimated disparities were incorrect there will remain disparities that can be corrected at the next step of the algorithm, using information from the next higher waveband in the images. Since at each layer, the two images are supposed to match more or less, thanks to the warping step, only a neighbourhood of five by five pixels is needed for each pixel in order to find the matching pixel in the other image.

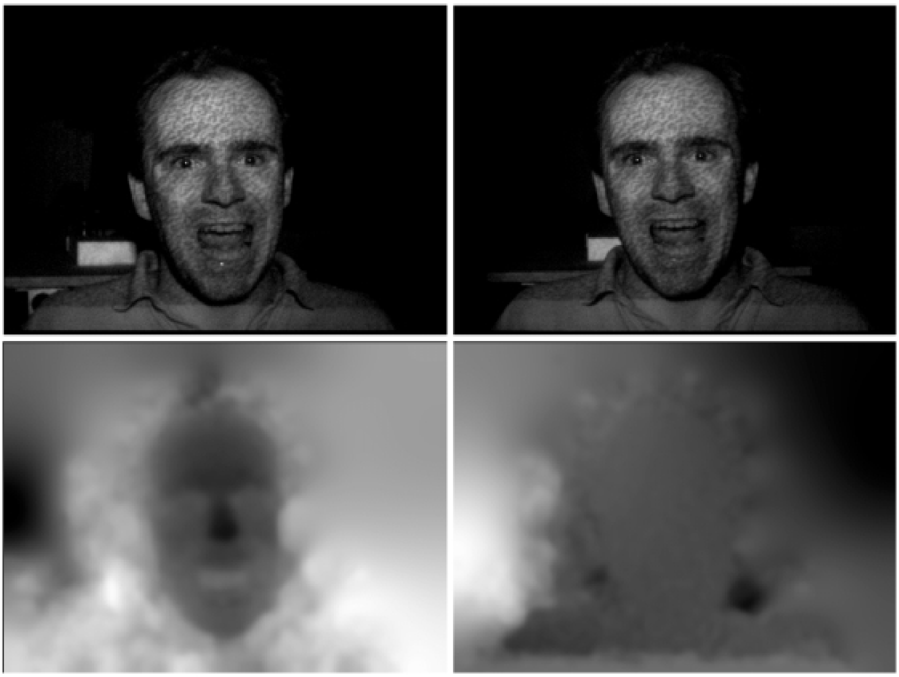


Fig. 2. Input images and final disparity maps (x,y)

Once the stereo matching process is completed, the final displacement files combined with the calibration file of the associated pod allow the generation of a range map, i.e. the map of the distances between each pixel and the coordinate system of the pod. Since the pods have been calibrated together, the 8 range maps of a given time step can be integrated in a single coordinate frame. A implicit surface is computed that merges together the point clouds into a single triangulated polygon mesh using a variant of the marching cubes algorithm [10]. This mesh is then further decimated to any arbitrary lower resolution for display purposes.

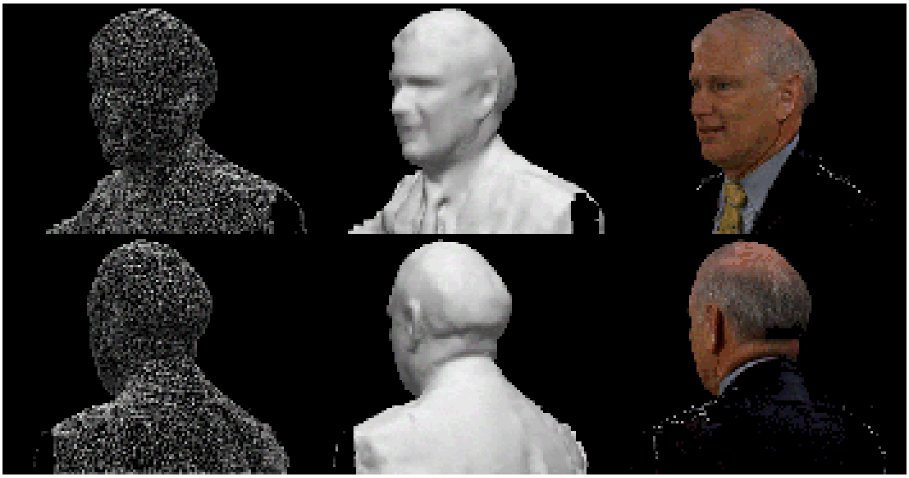


Fig. 3. Photo-realistic 3D model captured using 4 pods

The generation of photo-realistic models is achieved by mapping the colour pictures taken by the colour cameras to the 3D geometry. On Fig. 3, a photo-realistic 3D model, generated from four pods, is presented.

Imaging systems can offer a coverage of 90-95% of the full human body. Therefore the 3D models which are generated will not be complete, what is not acceptable for most applications. However we have recently worked on the conformation from generic models to 3D scanned data [7]. Consequently by conforming a complete generic model to our incomplete scanned models we can get an approximation for the missing pieces of mesh. Regarding the missing pieces of the texture, we will investigate the interpolation of the texture available and the utilisation of the texture generated at other time steps when there is no occlusion of the area of interest.

5 Results and Applications

5.1 True 3D Film

We present our first results, a true 3D film captured from four pods (head scanner). In total this film is composed of 25 frames (1 second), which represents 150 MB of raw data and 80 MB of VRML and JPEG files for a mesh resolution of 3mm. The data were processed fully automatically using a network of 4 PCs (PIII 803MHZ, 1GB RAM). The total computation time was of 38mn (25mn for the matching process and 13mn for the generation of the 25 3D models).

Fig. 4. 4 frames of a true 3D film captured by one pod

On the Fig. 4 we show the four first frames captured by one of the four pods and the corresponding 3D model generated from these pictures. The models are shown from different viewpoints. Obviously we do not think that pictures are the best supports for showing our results (the film can be downloaded from our web site [9]).

We think that this first true 3D film has achieved its goals. It demonstrates that our technology is reliable and up to the challenge. First, one second of a 3D film can be generated automatically in a bit more than half an hour. And secondly and more importantly the film is realistically believable, it looks like a real film instead of a great achievement of computer graphics.

Obviously the main limitation of our 3D studio is that, since there will be only 8 pods, it will not allow the capture of more than one actor at a time. However that should not prevent the generation of film involving several actors. Our 3D studio could be used as a 3D "blue screen" studio, where actors would be filmed separately in 3D. Then their 3D models could be integrated in any 3D environment (realistic or not) using classical 3D graphics techniques. In the future, the 3D studio could be fitted by much more cameras which would allow the generation of 3D films with several actors: the Robotics Institute of Carnegie Mellon University has demonstrated that using 49 cameras placed at different viewpoints, it is possible to capture two characters interacting with each other (see [14]).

5.2 Applications for Virtual Storytelling

The technology we have been developing gives the opportunity of telling virtual stories with many new ways. We could classify these applications in two closely related and often mixed categories: virtual storytelling based on virtual reality and animation techniques and virtual storytelling based on cinema and special effects.

Since the animation of virtual actors is still a very difficult task, real actors could be filmed in 3D using our studio and then these models would be integrated in virtual environments. Obviously these data could be edited and modified. Some interesting opportunities come from the fact that as we generate models using a marching cubes algorithm, a full sequence may be defined as a collection of 4D voxels associated to texture maps. Therefore when a model is built for a specific moment of the virtual story, it is possible to do it combining voxels created at different time steps. For example, we could visualise easily a scene where the speed of light would have been slowed down to few metres per second: the position of the extremities of the different limbs of a character would be few frames late compared to the position of the centre of the body.

In the near future we think that true 3D films will be used mainly at the production level in order to generate films that could not have been generated without our technology. At first, stories could be told by setting the position of the camera without any physical restriction. For example the films could be watched from the viewpoint of any character, human or not: we could imagine a second release of the successful 1999 fantasy comedy "Being John Malkovich", where the film would be shown from the eyes of John Malkovich! Secondly, stories could be also told with different sensitivities according to which pre-set path of viewpoints is chosen: the focus could be set on different characters such as the victim or the criminal, or a film could be suitable for an adult public under a specific viewing path and suitable for a younger public under a different one. In few years we should be able to offer to the public a unique experience: a full 3D immersion inside a film. By projecting the film using polarized light stereoscopy, the spectator, equipped with glasses and a joystick, will have the possibility of watching the film in 3D from any viewpoint they will navigate to.

Our technology is only at its prototype phase, but already many applications have been foreseen. There is no doubt that we will have to wait for its development and the involvement of more creative people in order to have a better idea of its full potential.

6 Conclusion and Future Work

In this paper, after having presented the state of the art in the fields of creation and animation of virtual humans, we described the 3D studio we are currently developing. Then we demonstrated the validity of the concept by generating a true 3D film using a system configured for head scanning. Finally we offered many applications of this technology for virtual storytelling.

In the future we will complete the system and optimise the computation, specially the matching process which is the most time consuming. Moreover since the amount of data that our 3D studio generates is quite important, we will need to address, as well, the compression of our data and new representations for 3D models.

References

1. V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH'99 Conference Proceedings, pp 187-194*, 1999.
2. M. F. Cohen C. Rose and B. Bodenheimer. Verbs and adverbs: Multidimensional motion interpolation. *IEEE Comp. Graphics and application*, 18(5):32–40, 1998.
3. D. J. Densley and P. J. Willis. Emotional posturing: A method towards achieving emotional figure animation. In *Computer Animation 97*, pages 8–14, 1997.
4. S. Etienne. *Positioning articulated figures.* PhD thesis, University of Glasgow, 1998.
5. S. Guo and J. Roberge. A high-level control mechanism for human locomotion based on parametric frame space interpolation. In *Computer animation and simulation'96*, pages 95–107. Springer Computer Science, 1996.
6. A. Hilton, D. Beresford, T. Gentils, R. Smith, and W. Sun. Virtual people: Capturing human models to populate virtual worlds. In *CA'99 Conference, 26-29 May 1999, Geneva, Switzerland*, 1999.
7. X. Ju and J. P. Siebert. Conformation from generic animatable models to 3d scanned data. In *International conference of numerisation 3D - Scanning*, 2001.
8. Takeo Kanade, Hideo Saito, and Sundar Vedula. The 3d room: Digitizing time-varying 3d events by synchronized multiple video streams. Technical Report CMU-RI-TR-98-34, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, December 1998.
9. 3D-MATIC Research Lab. http://www.dcs.gla.ac.uk/ jc/films/theawakening.html.
10. W. E. Lorensen and H.e Cline. Marching cubes: a high resolution 3d surface construction algorithm. In *ACM Computer Graphics, 21*, 1987.
11. T. P. Monks. Measuring the shape of time-varying objecs. In *PhD thesis, University of Southampton*, 1994.
12. J.-C. Nebel. Realistic collision avoidance of upper limbs based on neuroscience models. In *Computer Graphics Forum, volume 19(3), August 2000*, 2000.

13. J.-C. Nebel, F. J. Rodriguez-Miguel, and W. P. Cockshott. Stroboscopic stereo rangefinder. In *3DIM2001: Third International Conference on 3D Digital Imaging and Modeling*, 2001.
14. Virtualized Reality. http://www.cs.cmu.edu/ virtualized-reality/.
15. J. P. Siebert and S. J. Marshall. Human body 3d imaging by speckle texture projection photogrammetry. In *Sensor Review, 20(3),pp 218-226*, 2000.
16. TCTi. http://www.tcti.com/.
17. R. Trieb. 3d-body scanning for mass customized products - solutions and applications. In *International conference of numerisation 3D - Scanning*, 2000.
18. G. Vareille. Full body 3d digitizer. In *International conference of numerisation 3D - Scanning*, 2000.
19. Sundar Vedula, Simon Baker, Peter Rander, Robert Collins, and Takeo Kanade. Three-dimensional scene flow. In *Proceedings of the 7th International Conference on Computer Vision*, September 1999.
20. D. J. Wiley and J.K. Hahn. Interpolation synthesis of articulated figure motion. *IEEE Computer Graphics and application*, 17(6):39–45, 1997.
21. S. Winsborough. An insight into the design, manufacture and practical use of a 3d-body scanning system. In *International conference of numerisation 3D - Scanning*, 2000.
22. J. Zhengping. On the multi-scale iconic representation for low-level computer vision systems. In *PhD thesis, The Turing Institute and The University of Strathclyde*, 1988.

Spatial Sound Enhancing Virtual Story Telling

David Murphy and Ian Pitt

Computer Science Department,
University College, Cork, Ireland.
d.murphy@cs.ucc.ie and i.pitt@cs.ucc.ie

Abstract. Spatial sound information/cues may enhance the sense of immersiveness in virtual story telling. However, their role within complex, loosely-structured narratives is little understood. This paper describes a virtual heritage project that aims to convey a factual story using interactive virtual environments. Sound was added to the existing project in an effort to enhance the virtual experience. The use of sound is assessed through a user-study in order to assess its effectiveness and suggest methods of improvement.

1 Introduction

This paper concerns the use of sound in a virtual recreation of the North Main Street area of Cork City, Ireland. North Main Street was once the main thoroughfare of the city, and the model (built in VRML97) is designed to be used by visitors to the North Gate Centre, a visitor resource centre located in the heart of the North Gate area. It is also hoped that the model will be used by local schools. The aim is to stimulate interest in the history of the area among both casual and scholarly visitors, and thus to encourage exploration of the medieval remains in the area.

The model was built by a group of Multimedia Technology students working in collaboration with staff at the North Gate Centre [1]. It shows the area both as it was in the 17th century and today, and allows users to switch at will from one century to another in order to see how buildings, living conditions, etc., have changed over time. In addition, it includes links to movie segments and textual narratives that describe the history of the area and the many individual stories that contribute to this history.

1.1 Description of the Model

The model contains three virtual views of the city:

* An overview of the complete city centre as it was in the 17th Century. This model is designed to place the North Gate area in context. It is primarily used as a 'flyover' world, and has relatively little detail
* A representation of the North Gate area as it was in the 17th Century. This model has much greater detail than the 'flyover' world and is intended to be walked through rather than flown over.
* A representation of the North Gate area as it is today. Again, this model has much greater detail than the 'flyover' world and is intended to be walked through.

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 20–29, 2001.

In addition, a virtual museum was created, containing 3D models of artefacts known or thought to have been used in Cork in the 17th century.

The overview concentrates on the area enclosed within the old city walls and was based on 17th century maps and illustrations of Cork. A reasonable amount of information was available concerning the buildings within the city. Complete authenticity would have demanded that each building be created separately, but this would have increased the complexity of the model considerably. Therefore the buildings were categorised several basic types which were reused as necessary.

The models of the North Gate area in the 17th century and 21st centuries cover a much smaller area than the ’flyover’ city centre model, but are much more detailed. Whereas the houses in the city centre model are constructed from simple geometrical shapes and are solid, those in the North Gate models are much more finely-detailed and can be explored internally. The interiors are textured and, in most cases, contain furniture and other features. Some buildings were created in even greater detail, e.g., Skiddy’s Castle and Christchurch, which are the two most obvious landmarks within the area as well as being of exceptional interest as buildings. Some elements of the North Gate models are animated, e.g., an inn-sign in the 17th century model swings as if in a slight breeze.

Fig. 1. Models of the North Gate Area (a) in the 17th Century and (b) in the 21st Century

The ’flyover’ tour of the complete city centre forms the starting point for exploring the models. The user is taken on a pre-defined tour that takes in 21 viewpoints. Within the flyover tour, the user is always presented with a Head-Up Display (HUD) that remains in the same position on screen regardless of orientation and contains a set of control buttons. These allow the user to:

* Start or stop the tour
* Move to the more detailed model of the North Gate area in the 17th Century
* Go to the virtual museum.

Access to the 20th century walkthrough model of the North Gate area is via links from the 17th century model of the North Gate area. Many of the 17th century buildings contain ’hotspots’ that, when activated, cause the 17th century building (and its immediate surroundings) to disappear, and be replaced by whatever is on the same site today. In this way the 17th century model can be replaced with the 21st century model section-by-section. The process is reversible, so that when some or all of the

21st century model is visible the user can click-on hotspots to see what occupied each site in the 17th century. Alternately, an always-visible button allows the user to move directly between the 17th and 21st century models, retaining the same viewpoint.

1.2 The Need for Sound

Upon completion, the project was used by a number of groups and individuals. It was observed that most users visited all the models and enjoyed using the 'hotspots' to jump from one to another, but that few spent much time exploring the other features of the project. In this sense the project fell short of its aims: the level of detail and amount of information available limited its appeal to scholarly users, while other users appeared to be more interested in its interactive features than its content. This was particularly true of children, one of the groups at whom the project was aimed.

One possible solution was to add sound to the models. The original project used sound in several places - principally voice-overs and sound effects for movie segments - but not within the models themselves. Sound and in particular spatial sound is a very powerful medium for conveying a sense of immersion within a virtual environment. The various attributes of spatial sound can affect the perception of an environment, the context of a communication, and alter the meaning of a message. Therefore it seemed likely that adding sound might increase the user's feeling of involvement without greatly increasing rendering times, etc..

In view of this, the authors decided to use sound in a number of ways:

Foley effects and **earcons** are incorporated to enhance the realism of objects within the environment. For example, the inn sign that swings in the wind (see above) is accompanied by a creaking sound that varies in synchrony with the movement. Similarly, when a door opens, an associated creaking sound is played in synchrony with the movement of the door.

Environmental sounds and music: for example, when the visitor enters an old tavern, spatial sounds reminiscent of the period and of the environment are rendered, such as general chatter augmented at intervals by the sound of drink being poured, and the chinking of glasses. Period music is also used.

Environmental sounds are often used to give a feeling of immersion in an environment, but in this case they were also used as a 'lure' to persuade users to explore further. Whilst in a particular environment, users hear the sound associated with that environment plus - in the distance, and with suitable localisation cues - the sounds associated with adjacent environments. For example, whilst exploring a room in Skiddy's Castle, the user might be able to hear the sounds of pots and pans being chinked in the nearby kitchen, or a feast taking place in the Great Hall. These sounds are normally heard at a low volume relative to that of the sound associated with the immediate environment, but periodically they rise in volume briefly so that they are clearly audible above the local sounds, thus ensuring that they are heard.

No attempt was made to simulate the events associated with these sounds. For example, although the user can hear pots and pans being used in the castle kitchen, the kitchen itself is empty. Thus the sounds seem to be produced by ghosts, who are heard but not seen.

2 Sound in Virtual Displays

Creating sounds with suitable spatialisation in VRML97 presents a number of challenges, as does controlling them realistically at run-time.

2.1 Sound in VRML

The Sound Node in VRML contains ten fields (for more information see [2] and [3]). Four of these ('minFront', 'minBack', 'maxFront' and 'maxBack') define the radiation pattern of a sound object. This pattern is restricted to an elliptical shape and, consequently, sound objects are treated as directional sounds. The direction of the sound, which corresponds to the apex of the ellipsoid (see Figure 2), specifies the path along which the direct sound will travel. This vector is specified in the 'direction' field.

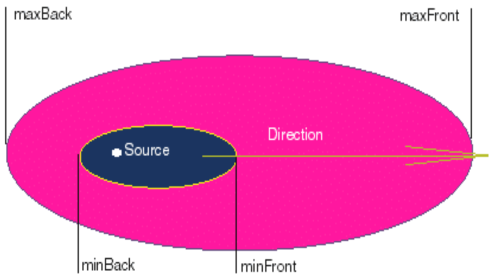


Fig. 2. Elliptical Model for Sound Source

The sound source is bounded by two ellipsoids within which a linear attenuation is performed. The inner ellipsoid has an intensity of 1.0 (the maximum level, 0dB) and the minimum level of the sound 0.0 (-20dB) is determined by the outer ellipsoid. However, the cut-off at -20dB is abrupt and very audible. The outer ellipsoid also behaves as an acoustic proximity sensor and when the user traverses this boundary the sound is activated.

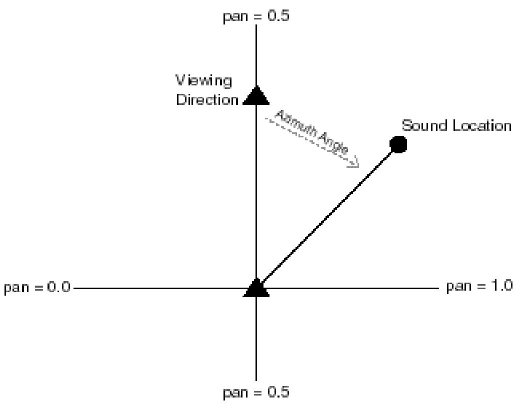


Fig. 3. VRML Stereo Panning Model

The location of the sound source is specified in the 'location' field; this is a 3D coordinate position given in the X, Y and Z-axis. The final attribute, which activates the spatialisation of the sound object, is the 'spatial' field. The 'spatial' field is a Boolean field and if set to TRUE will activate the spatial mechanism on the end-user's system. As a minimum the VRML standard specifies that "Browsers shall at least support stereo panning of non-MIDI sounds based on the angle between the viewer and the source" [4]. For simple stereo panning, the location of the source is mapped to the X and Z planes to determine the azimuth of the source in relation to the viewer (Figure 3). This angle is then assigned a pan value between 0.0 and 1.0. However, it is recommended that the Browser should use a more sophisticated technique of spatialisation than basic amplitude panning [4]. This technique is a restricted form of spatial sound rendering and not very immersive.

Another shortcoming of sound spatialisation in VRML is that there is no explicit use of height/elevation information or even recommendations for its use. Height information could be derived from a comparison of the listener position/orientation and the sound source location on the Y-axis.

2.2 MPEG-4 Sound

MPEG (Motion Picture Experts Group) is a working group of an ISO/IEC subcommittee that creates multimedia standards. In particular, MPEG defines the syntax of low bitrate video and audio bit streams, and the operation of codecs. MPEG has been working for a number of years on the design of a complete multimedia toolkit, which can generate platform independent, dynamic interactive media representations. This went on to become the MPEG-4 standard.

In this standard, the various media are encoded separately; this allows for better compression, the inclusion of behavioral characteristics and also enables user-level interaction. Instead of creating a new scene description language the MPEG organization decided to incorporate VRML.

As noted earlier, VRML's scene description capabilities are not very sophisticated so MPEG extended the functionality of the existing VRML nodes and incorporated new nodes with advanced features. Support for advanced sound within the scene graph was one of the areas developed further by MPEG.

The Sound Node[1] of MPEG-4 is quite similar to that of the VRML Sound Node. However, MPEG-4 contains a sound spatialisation paradigm called 'Environmental Spatialisation of Audio' (ESA). ESA can be divided into a Physical Model and a Perceptual Model.

Physical Model: This enables the rendering of source directivity, detailed room acoustics and acoustic properties for geometrical objects (walls, furniture, etc.). 'Auralisation', another term for the physical model, has been defined as:

> "creating a virtual auditory environment that models an existent or non-existent space." [5]

Three Nodes have been devised to facilitate the physical approach. These are AcousticScene, AcousticMaterial and DirectiveSound.

---

[1] A detailed listing of the MPEG-4 and VRML Sound Nodes can be found in [2]

Briefly, DirectiveSound is a replacement for the simpler Sound Node. It defines a directional sound source whose attenuation can be described in terms of distance and air absorption. The direction of the source is not limited to a directional vector or a particular geometrical shape.

The velocity of the sound can be controlled via the 'speedOfSound' field; this can be used, for example, to create an instance of the well-known Doppler Effect. Attenuation over the 'distance' field can now drop to -60dB and can be frequency-dependent if the 'useAirabs' field is set to TRUE.

The 'spatialize' field behaves the same as its counterpart in the Sound Node but with the addition that any reflections associated with this source are also spatially rendered. The 'roomEffect' field controls the enabling of ESA and if TRUE the source is spatialized according to the environment's acoustic parameters.

AcousticScene is a node for generating the acoustic properties of an environment. It simply establishes the volume and size of the environment and assigns it a reverberation time. The auralisation of the environment involves the processing of information from the AcousticScene and the acoustic properties of surfaces as declared in AcousticMaterial.

Perceptual Model: Version 1 of the MPEG-4 standard only rendered spatial sound based upon physical attributes, i.e. geometric properties. However, virtual worlds are not constrained by physical laws and properties; therefore it was necessary to introduce a perceptual equivalence of the physical model. To this end, two new Nodes were added in version 2 of MPEG-4; PerceptualScene and PerceptualSound. Rault et al, highlighted the merits of the perceptual approach in a recent document to the MPEG group:

> "A first advantage we see in this concept is that both the design and the control of MPEG4 Scenes is more intuitive compared to the physical approach, and manipulating these parameters does not require any particular skills in Acoustics. A second advantage is that one can easily attribute individual acoustical properties for each sound present in a given virtual scene." [6]

The principle elements of the perceptual model are drawn from research undertaken by IRCAM's Spatialisateur project, and additional features are derived from Creative Lab's Environmental Audio Extensions (EAX) and Microsoft's DirectSound API [7]. Using the perceptual model, each sound source's spatial attributes can be manipulated individually, or an acoustic-preset can be designed for the environment.

Fields such as 'Presence', 'Brilliance', and 'Heavyness' are used to configure the room/object's acoustic characteristics. In all, there are nine fields used to describe, in non-technical terms, the spatial characteristics of a room or a sound object. These fields have been derived from psycho-acoustic experiments carried out at IRCAM (Spatialisateur Project). Of the nine subjective fields, six describe perceptual attributes of the environment, and three are perceived characteristics of the source. Table 1 lists the parameters for both Environment and Source.

It can also be seen from Table 1 that the last three fields of the Environment section and all of the Source fields are dependent upon the position, orientation and directivity of the source.

The validity of this approach could be questioned in terms of its subjectivity, for example, the choice of words such as 'Warmth' and 'Brilliance'. However, the use of subjective terms as acoustic parameters, in this context, is to facilitate the non-specialist to compose a soundscape with convincing acoustic properties. This effectively opens up the complex world of acoustics to the non-specialist.

Table 1. Perceptual Fields for MPEG-4 Spatial Audio

Environment Fields	Source Fields
LateReverberance	Presence
Heavyness	Warmth
Liveness	Brilliance
RoomPresence	
RunningReverberance	
RoomEnvelopment	

3 Implementation

The playback of the sounds in virtual reality can be either headphone or speaker based. For the purposes of this research headphone playback was chosen. This enables the designer to incorporate more complex sound objects whose subtleties will not be lost due to background noise, speaker cross-talk, etc. Another advantage to using headphone reproduction is that a simple head-tracking device can be attached and relay six degrees-of-freedom information to the virtual scene manager.

Head Tracking is an important tool in any dynamic virtual environment. Apart from the added sense of immersiveness and dynamic visual presentation it is also important in the spatial rendering of sound. In the natural world head movement is used to obtain a better sense of a sound's direction and position. According to Burgess "The lack of these [head-related] cues can make spatial sound difficult to use... This 'closed-loop' cue can be added to a spatial sound system through the use of a head-tracking device." [8] Recent research has shown that the use of Head Tracking reduces source position reversal (i.e. the impression that the sound originates behind rather than in front of the listener, or vice-versa) by as much as 2:1 [9]. There is also evidence that it assists in the externalisation of sources that would otherwise be located 'inside-the-head'.

Another area where head-tracking is helpful is in the simulation and control of the Doppler Effect and to resolve source-listener movement ambiguities.

3.1 Spatialisateur

As the original North Main Street project was written in VRML it was decided to continue with this scene description language, otherwise a substantial rewrite of the original would have been necessary. As noted earlier VRML does support sound and includes a basic spatial sound model, however it is generally accepted that it is too

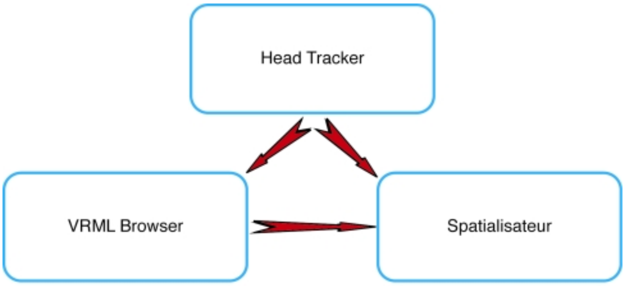


Fig. 4. Structure of the Spatial Sound Processing

basic for most interactive VR applications [2][3]. VRML is highly extensible and is well suited to proprietary extensions. Using this flexibility the authors designed a set of PROTO Sound Nodes (customised VRML Sound Nodes), which use an external tool to process the sounds and perform the spatial sound rendering.

Due to its high-level interface, efficient processing algorithms, and its ease of programming the authors chose IRCAM's Spatialisateur to process and spatially render the sounds for the North Main Street project. A high-level schematic (Figure 4) shows the relationship between the VRML Scene, Spatialisateur, and the user's position and orientation via the head-tracker.

Many of the features of Spatialisateur were incorporated into MPEG-4 and have already been explained, most notably the Perceptual Model. This approach combined with an intuitive high-level interface enables the swift development of a virtual auditory display.

Spatialisateur employs several binaural rendering techniques, of which the HRTF approach was determined to be the most appropriate for this research.

Distance and the perception of distance of sound sources within a virtual environment are controlled by the amplitude level of the sound and the amount of reverberation applied. When a sound is relatively close the amplitude is quite high and there is much more direct sound than there is reverberation. If, on the other hand, the source is distant the amplitude level will be low and there will be a large amount of reverberation. When these two factors are combined with HRTFs a truly immersive three-dimensional virtual audio display can be created - this is the technique adopted by the authors in the sonification of the North Main Street project.

4 Evaluation

Having added sounds to the model, the authors set out to determine how the inclusion and use of sounds affected users' exploration. It is intended that a number of studies will be carried out using the model. However, in the first instance, the authors set out to determine:

* how the use of sounds affected users' perceptions of the model (e.g, did they enhance it or not)
* if the use of sound encouraged greater exploration of the model (and the narratives within it).

To this end, a simple user study was conducted. Two groups of school-age children were invited to use the system: one group used the system in its original form, without sound, while the other group used the system with spatial sound added. The intention was to find out how long the system could hold the attention of a child, and to see if and how this varied with the inclusion of sound. To this end, the children were given no specific goals or tasks to complete. Each was given a brief introduction to the system, and then allowed to spend as much time using the system as he/she wished.

The principal variable measured was time spent using the system. The total time was recorded, and also the time spent within specific areas of the model. Thus it was possible to determine whether the inclusion of sound changed the total length of time spent exploring the model, the pattern of exploration, or both. In addition, the children were asked to rate the system for enjoyment.

This study is continuing, but early results suggest that the addition of spatial sound increases childrens' enjoyment and encourages greater exploration. On average, subjects spend longer using the audio-enhanced version of the model and claim to have derived greater enjoyment from using it.

The use of environmental sounds as a lure - e.g., the use of spatial sounds to indicate that there is something of interest outside the currently-visible environment - also appears to be effective in encouraging greater exploration, although it may have disadvantages as well as advantages. Subjects appear more willing to explore a room or space if it has associated sound. However, there is also some evidence that a subject may curtail exploration of the current area if the sounds suggest that an adjacent area may be more interesting.

The fact that the 'lure' sounds were not accompanied by any corresponding visible events did not appear to cause any problems (for example, hearing sounds of pots and pans being used in the kitchen, without seeing cooks and kitchen staff). The subjects seemed to accept that they were hearing the ghost-like echoes of events that had taken place in the past, and although some commented on this, few seemed disturbed by it.

5 Conclusions and Further Work

The study raises a number of issues.

First the results of the study suggest that the addition of sound encourages exploration of a virtual environment. In this case, the motivation for using sound was to try and encourage greater exploration of a virtual environment by children without otherwise changing the environment itself (e.g., by increasing visual realism or adding a game-like structure to the interaction). No attempt was made to compare the efficacy of sound with other approaches, but the results suggest that adding sound may in itself go a long way towards achieving this end.

Part of the reason for using sound is that it potentially offers a way to increase the sense of immersion in an environment without significantly increasing processing overheads. However, a drawback is that the approach described here relies on accurate localisation of sound sources, and this is difficult to achieve without expensive external hardware. Thus it is not necessarily the case that using sound can be seen as a cheap substitute for greater graphics processing power. However, it is possible that even a relatively simple audio implementation would yield some benefits. This would allow audio and graphics performance requirements to be balanced against one another when attempting to achieve a sense of immersion on relatively low-cost platforms.

The use of environmental sounds as a 'lure' to encourage exploration of areas outside the currently-visible range seems to work, but has limitations, e.g., the observation that users may curtail exploration of an area if the sounds suggest that an adjacent area may be more interesting. This issue needs to be explored further, but even if problems do exist they need not be regarded as insurmountable. It might be possible, for example, to delay the introduction of the 'lure' sounds until a user has spent a certain amount of time in a particular space, or shows signs of leaving the space.

The use of sound in this way also opens-up the possibility of using sound to make virtual worlds more accessible to blind and visually-impaired users. It has been suggested that blind users might be able to explore models such as the one used in this study with the aid of a tactile-feedback device (such as the Phantom). However, this would only allow the user to explore a single point at any one time. Sighted users would be able to take in much of the shape of (e.g.) a room at a single glance, and view the rest simply by panning around a fixed viewpoint. Thus they would be able to see items of potential interest and move towards them, so discovering new links. However, blind users would only be able to determine the shape of a room and find any links or objects it contains by exploring it fully. This would be much slower, and there is no guarantee that everything of interest would be found. Using sound, it might be possible to signal the existence and location of items of potential interest and thus give blind users some of the 'scanning' ability available to sighted users. It is intended that this issue will be explored in a further study.

References

1. O'Riordan. C., et al., 'Cork Through the Ages - A Multimedia History of the Walled City', Master's Thesis, University College, Cork, Ireland, 2000.
2. Murphy, D., 'Evaluation of Spatial Sound Description in Virtual Environments', Cambridge Music Processing Colloquium, Cambridge, September 1999.
3. Väänänen, R. and Huopaniemi, J., 'Spatial Presentation of Sound in Scene Description Languages', 17th AES International Conference, Munich, May 1999.
4. ISO/IEC JTC/SC24 IS 14772-1 "The Virtual Reality Modeling Language (VRML97) Information technology - Computer graphics and image processing - The Virtual Reality Modeling Language (VRML)". URL: http://www.vrml.org/Specifications/VRML97/.April1997
5. Väänänen, R., 'Verification Model of Advanced BIFS (Systems VM 4.0 subpart 2)', ISO/IEC JTC1/SC29/WG11 Document N2525
6. Rault, J-B. et al., 'Audio Rendering of Virtual Room Acoustics and Perceptual Description of the Auditory Scene', ISO/IEC JTC1/SC29/WG11 Document M4222
7. Jot, J-M., Ray, L. and Dahl, L., 'Extensions of Audio BIFS: Interfaces and Models Integrating Geometrical and Perceptual Paradigms for the Environmental Spatialisation of Audio', ISO/IEC JTC1/SC29/WG11 Document M4223
8. Burgess, D., 'Techniques for Low Cost Spatial Audio', 1994, Located at ftp://ftp.ircam.fr
9. Begault, R. et al.. 'Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source', AES, 108th Convention Paris, France, February, 2000

The VISIONS Project

Olivier Balet[1], Paul Kafno[2], Fred Jordan[3], and Tony Polichroniadis[4]

[1] CS SI, Virtual Reality Department, ZAC de la Grande Plaine,
rue Brindejonc des Moulinais, 31500 Toulouse, France
Olivier.Balet@c-s.fr
[2] HD Thames, 199 Piccadilly, W1V 9LE London, UK
hdthames@easynet.co.uk
[3] ZVisuel, Avenue du Léman 40, CH-1005 Lausanne
Fred.Jordan@zvisuel.com
[4] Anthropics Technology Ltd, Ealing Studios, Ealing, London W5 5EP, UK
tony@anthropics.com

Abstract. This paper presents VISIONS, an innovative software product using the latest Virtual Reality technologies for the virtual prototyping and authoring of various forms of stories. VISIONS is developed in the frame of an IST project funded jointly by the European Commission Fifth Framework RTD Programme and the VISIONS consortium.

1 Introduction

The Story is one of the riches of western civilisation, and its many forms lie at the heart of the world's most important industry. Feature films, cartoons, television, advertising and multimedia works are all built around fundamental narrative structures. Virtual Reality now offers the excitement of manipulating these forms in new, non-linear ways while providing means to enhance their production.

The VISIONS project aims at developing an innovative software product using the latest Virtual Reality technologies for the virtual prototyping and authoring of various forms of stories. This two-year long project has been initiated in January 2000 by a consortium of 6 partners leaded by CS SI, the French leader in Virtual Reality software development. The five other partners are Anthropics, a British company expert in character animation, ZVisuel, a Swiss company expert in 3D interactive object technologies, Giunti Multimedia, Italy's oldest publisher and multimedia producer, Nord-Ouest, a French feature and advertising film company, and HD Thames, an independent television production company.

The VISIONS system, developed in the frame of the project, has been designed to encourage authors to conceive stories visually, and then enable them to output a dynamic and persuasive virtual mock-up of their work. It really helps any kind of author, director or producer to go beyond the illustrative capabilities of traditional storyboards, scenarios and scripts.

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 30–39, 2001.

VISIONS can output data under different formats: as a digital video for fund raising and marketing phase, as storyboards, scripts and sketches for the production team or as a 3D interactive story for electronic publication.

The key issue of the project is to develop a user-friendly system that could be efficiently used by people with no computer skills on mainstream PC platforms. This is an extremely challenging objective because of the inherent three-dimensionality of VISIONS. It is indeed demonstrated that working with three-dimensional software is very difficult [6] and still requires high skills, hours of practice and training. Even the simplest navigation task within a basic 3D environment can turn into a nightmare for most of the users. It is even worst when one has to go beyond simple navigation and needs to manipulate 3D objects. In this domain, the animation of 3D articulated figures, such as virtual actors, requires the ultimate in computer graphics skills. The lack of correlation between manipulation and effect and the high cognitive distance from users to visualized 3D models on 2D screens are the major reasons for these problems.

In the following sections, we present the technological solutions we have chosen, developed and combined to obtain a 3D application as user-friendly, intuitive and efficient as possible without requiring any particular computer skill from the user. Similar approaches [2], [3], [15] have been recently successfully adopted for developing highly intuitive systems for industrial virtual prototyping.

2 Architecture

The VISIONS software architecture has been designed to be opened and modular in order to allow both the development of external components or plug-ins. It is composed of 5 main components developed in C++: the VISIONS Core, the VISIONS SDK, the VISIONS Application, a set of two plug-ins for casting and directing virtual actors, the 3D Living Model Collection.

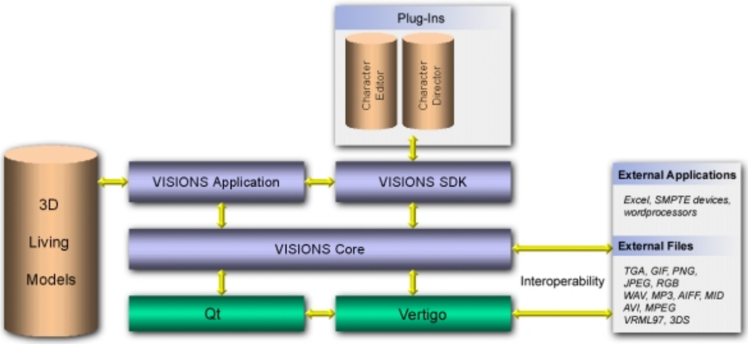


Fig. 1. Overall Software Architecture of the VISIONS System

The VISIONS core. It has been built on top of Qt for the 2D GUI aspects, and Vertigo for the 3D interaction and rendering aspects.

Troll Tech's Qt™ toolkit has been selected for developing the Graphical User Interface. It is a European multi-platform GUI toolkit that has been widely and successfully adopted by the Linux community (KDE is based on Qt).

Vertigo is CS' set of C++ libraries for developing cross-platform VR and real-time 3D applications. It is fast and integrates many VR device drivers (3D controllers, CrystalEyes VR, 3D sensors, etc.), file format import/export modules (including VRML97 and 3DS) and state-of-the-art 3D rendering techniques (lens flares, particle systems such as rain, snow or smoke, per-pixel lighting, BSP, etc.).

The VISIONS core provides all the basic modules for: interoperating with external software or importing/exporting files under various formats, managing natural and multi-modal man-machine interaction (input and output), developing the application, the SDK and external plug-ins. Several import/export modules have been developed in order to allow users to integrate existing content (images, sounds, 3D models) into their story or to communicate with external software such as Word, Premiere, etc.

The VISIONS SDK. Its role is to give technology partners and third-party developers the tools needed to develop new plug-ins for VISIONS. The SDK gives developers full access to the 3D layer, the Graphical User Interface, the story database containing all the story elements (actors, props, lights, sets, soundtrack, actions, shots, etc.), and a set of file management tools.

The VISIONS application. This module is made up of a very light kernel and a set of standard plug-in modules. This kernel is responsible for creating, loading and saving a project, customising the interface and managing the plug-ins. The standard plug-ins provide the basic functionalities such as the tools for creating, visualising and editing the sequences, the timeline, the cameras and the sets.

In the following, we briefly present 4 of the main standard plug-ins that compose the VISIONS application.

The Timeline Window Plug-In. The timeline window is the 2D representation of the scenario for a specific sequence, a sequence being defined by all the events that occur in a set during a certain time. These events are hierarchically represented. Vertically, the timeline is divided into tracks grouped by element. For instance, one group of tracks is used for the events (movements, dialogue, and actions) related to a particular actor, while another group is used for describing the trajectory of his car or the behaviour of his mobile phone. The timeline allows the user to easily place the events in time, to modify their duration and edit their content.

Fig. 2. The Timeline Window

The timeline clock is based on a SMPTE time code thus allowing the synchronisation of the system with external devices (MIDI sequencers, video recorders, etc.) or software (Cubase, Cakewalk, etc.) thanks to SMPTE and MTC/MMC synchronisation protocols.

The Camera Window Plug-In. The camera windows provide 3D rendering of the selected sequence. Selection, edition and manipulation of the elements are directly achievable through their 3D representation. Several overlaying indicators such as rulers, grids or ground glasses can be displayed to make the 3D camera operation easier for the user.

In addition, it is possible to use one or two joysticks to operate the VISIONS cameras, thus mimicking the camera controller used to operate a real camera head.

Fig. 3. The Camera Window

The Set Editor Plug-In. The set editor allows the user to build and dress outdoor and indoor sets. It provides a list of tools allowing the creation, decoration, manipulation, and edition of the set elements. It is possible to create elements from scratch, such as geometric primitives, terrains, floors, walls, doors, windows, stairs, etc., or to use the elements available in the VISIONS 3D Living Model Collection. Several creation or decoration metaphors, such as drag-and-drop, magic wand, paint brush, pickaxe, are always available to the user to provide him with an extremely flexible interface.

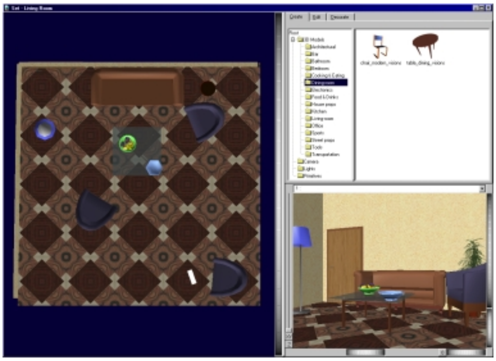


Fig. 4. The Set Editor

The 3D Manipulation Plug-In. VISIONS provides a set of 3D metaphors, called manipulators [6] [14], that allow the direct manipulation of the story elements. A 3D metaphor explains, in a pictorial and easy-to-understand way, the handling of a 3D operation with a 2-dimensional input device. A manipulator is made of different parts, each part is associated to a number of parameters. Clicking on a specific part and dragging it will affect the associated object according to the part movement (the green parts control translations, the red parts control rotations, the yellow parts control scales).

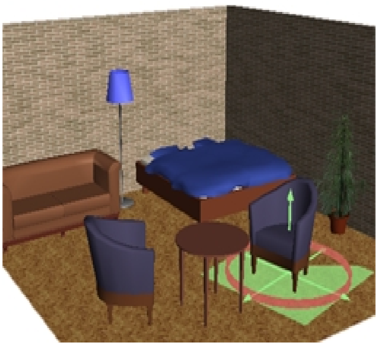


Fig. 5. A 3D Manipulator

3 Virtual Actors

Humans are perhaps the most complex single entities to have ever existed, and as a result, accurately modelling and animating people will always be a difficult and yet incredibly important part of computer graphics. This section introduces the concept of a Virtual Actor, a computer representation of a human character, as used in the VISIONS project. It outlines how VISIONS innovates in this domain by bringing enhanced video game and industrial 3D technologies to the world of story authors for the fast, easy, versatile and real-time casting and direction of virtual actors.

3.1 Virtual Actor Casting

Fig. 6. Quality obtained with less than 1000 polygons per character (models from geo-metricks.com)

Nowadays, creating the correct actors for a particular story without involving modelling experts is an extremely difficult task. This usually requires 3D modelling software (Maya, Softimage, 3DS-Max, etc.) that is not usable by non-expert users. The VISIONS Character Casting plug-in has been designed to allow any kind of user to intuitively "cast" characters by setting attributes, which can either be a selection from a list (such as "male" or "female") or numerical values (such as "nose size", "weight"), and costumes.

A character is defined by its skeleton, its outer shape layer, its face and its costume (including accessories). It is important to note that the system does not simply provide a selection of characters that the user can choose between.

Body Casting. The VISIONS body-casting module is based on a Free Form Deformation (FFD) algorithm that allows the user to intuitively modify an actor's morphology. FFD is a deformation technique, popularised by [13], that is independent of the object representation. Deformation of an object is done indirectly by deforming the space in which the object is embedded with a set of control points that act as magnets.

VISIONS hides the concept of control points behind a simple interface that allows the user to modify the actor body with sliders representing understandable attributes (ex: leg length, belly volume, breast firmness, etc.). Internally, the system uses two control point sets for each attribute. The first set defines the control point positions at a minimal value (ex: smallest leg) while the second set defines the control point positions at a maximal value (ex: longest leg). Thus, it is possible to automatically interpolate the position for each control point from a slider value and two control point sets.

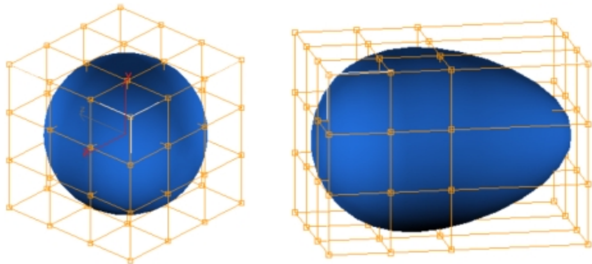


Fig. 7. FFD Control points

Fig. 8. Character Casting

Face Casting. A large number of 3D models of faces have been captured using a multi camera, vision based, 3D capture rig [5]. These faces have been processed with a statistical technique, the *Linear Discriminant Analysis* [7], that reduces their dimensionality and extracts abstract features. Thanks to this technique, we have obtained an extremely compact set of parameters that fully determines most of the features of the human face. VISIONS uses this set to reconstruct any new face from the indications (ex: European or African, young or old, fat or thin, blue or green eyes) provided by the user with the face editor window.

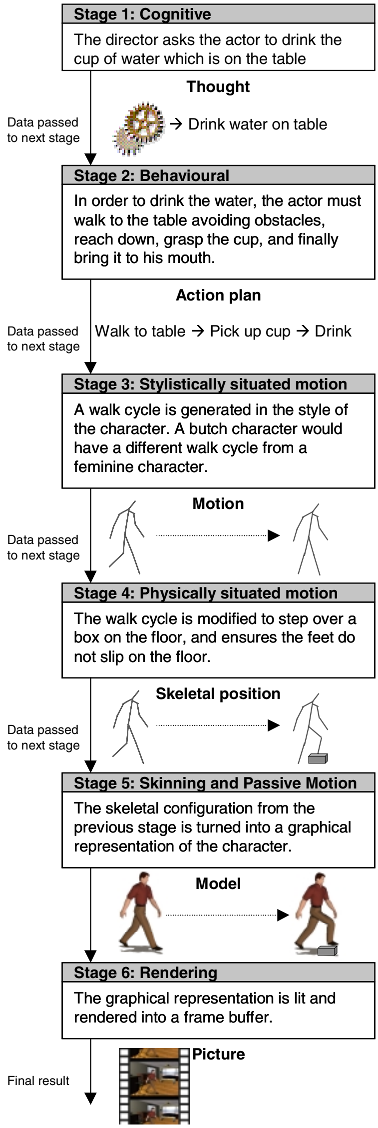
3.2 Virtual Actor Direction

The task of producing a complete, believable model of a person is obviously a complex one, but as with all complex tasks, breaking it down into simpler sub-tasks can bring great benefits. Morawetz [9], Perlin [11], Blumberg [4] and Badler [1] all describe complete models of a virtual actor system.

Although their techniques differ in implementation, they all use a similar model for breaking up the task, namely, selecting an action, producing movement from the action and rendering frames. We break up the task in a similar manner, but are finer grained.

Our conceptual model for producing Virtual Actors who obey the director is to use a pipeline. The process is modelled as a single directional flow of information through a set of processing units, similar in style to the graphics pipeline used in computer graphics. In fact, the last stage in the Virtual Actor pipeline is the graphics pipeline, which is used to render the graphical representation of the character.

The Virtual Actor Animation Pipeline. The first stage in the pipeline is the Cognitive Level. In order to achieve the appearance of being a motivated character, the Virtual Actor needs to be able to act on, and produce those motivations. Once a character has made a conscious decision to achieve something, such as to change location, they must determine what actions must be carried out to achieve that aim.



Once a set of atomic actions has been decided on, they must be turned into body motion. This motion must be consistent with the character they play.

Once a piece of motion has been created it is important that it looks correct for its physical context. This means that it cannot break any physical laws, and that it achieves all physical goals that are required of it.

The Skinning and Passive Motion stage takes the posed skeleton produced by the physically situated motion stage, and produces a graphical representation of it. This can involve many different elements, depending on the complexity of the visual representation.

Rendering is the final stage in the Virtual Actor pipeline. Here the surface information is drawn to produce the resulting picture.

The plot of the story determines the characters intentions and actions. The user can then select an action from a list, possibly done in different styles for the character to perform. These actions are performed by real people, and recorded onto the computer using motion capture.

Physically Situated Motion. In the case of motion captured from real life, the exact placements of objects and the situation in the motion capture studio most probably do not mirror the positioning of objects in the virtual set. The motion may also have been captured for a character of different dimensions than the intended character. For these two reasons, it is possible for motion to become divorced from the physical situation being created.

This may have the result that characters appear to reach through objects, or perform actions that are not physically possible in the virtual set. As people are used to natural motion, it is very easy to spot features of synthetic motion that break these rules. With current animation packages, the user must solve these problems manually, sometime spending hours to adapt the character's movements.

VISIONS uses a motion retargeting algorithm [8] [12] to automatically adapt prerecorded motions from a real actor to a virtual one who has a different morphology. A new algorithm [12] is also used to automatically adapt the virtual actor's motion sequence to his environment in order, for instance, to modify the original path to avoid collisions with the set elements.

4 The 3D Living Model Collection

An important part of the project has been devoted to the design and the creation of the 3D Living Model Collection (3DLMC). Like other 3D model libraries, it is composed of several hundred 3D textured objects that can be used as basic bricks for constructing sets and populating virtual stories. However, our library integrates two new concepts we propose for addressing the highly complex problem of the user/virtual environment interaction and the even more complex problem of the virtual actor/virtual environment autonomous interaction.

We indeed propose to embed within each active object of the library, a set of behaviours as well as a software model of the object instructions of use that will allow the system to simplify the interaction and animation procedures. And that is how the VISIONS system can allow a user to give a "grasp the bottle and drink it" order to hivirtual actor.

Background. The autonomous interaction between virtual actors and objects is probably one of the most complex problems of the computer graphics world. Some models have already been published for specific body parts such as the hand. However, no solution has yet been proposed for providing a unique mathematical model for this kind of interaction, that is a solution compliant with any body parts that could apply to any objects. The particular case of the hand-object interaction is described in [9]. This approach is very detailed but specific to the hand morphology and limited to grasping. [16] proposes a solution to describe some of the key features of each object in a way that is independent of the interacting body part. However, this description is different for each object and cannot therefore be used to automate the virtual actor/virtual object interaction. The following sections outline the generic solution we propose for supporting the automatic interaction between virtual actors and their environment

Abilities and Effectors. In the VISIONS application, the virtual actor can interact with an object in a pre-defined set of ways that are called **abilities** and are embedded in the object definition file. The set of abilities of an object defines its instructions of use. Each ability is related to a part of the virtual actor anatomy that is used for the interaction. This part is called the **effector**. For instance a bottle has two abilities (it can be held and drunk) and two corresponding effectors (the hand and the mouth).

The VISIONS Node. The file format used for the 3DLMC is currently VRML97. Data related to abilities are added to the VRML definition of the object in a script node, the VISIONS node. Therefore, VRML files featuring the VISIONS node stay fully compliant with the VRML specifications. Another approach [16] proposes to store the data in an `exposedField` of a VRML PROTO. The advantage of our solution is that it keeps the original structure of the VRML file.

Indeed a script node just needs to be appended to the original file whereas the use of PROTO requires an instantiation in order to “encapsulate” the object.

For each ability Ai (ex: "carryable"), this node defines the position Pi and the orientation Oi of the corresponding effector (ex: "right hand") in respect of the object through the following fields:

```
<abilities>        A1=ability1 ,…, An=abilityn
<positions>        P1(x1 y1 z1), …, Pn(xn yn zn)
<orientations>     O1(x1 y1 z1 a1), …, On(xn yn zn an)
```

where (xi yi zi) denotes the Cartesian coordinates of the position and (xi yi zi ai) denotes a rotation of the angle ai along the axis defined by the coordinates (xi yi zi).

Moving and Still Effectors. Effectors are parts of the body that can interact with an object. Two sets of effectors have been identified: the hand effector and the other parts of the body. The fundamental reason for making such a differentiation is the way the hand and other parts of the body interact with objects. Indeed, it is both the hand orientation and position of an actor that are modified in order to realistically grasp or hold an object. However, it is the object position and its orientation that are modified in order to fit with the head position for instance. This can be illustrated with the phone case. When a hand grasps a phone handset (this is the carryable ability), the hand must take a particular orientation and position but the handset itself keeps still. If we now consider the case of a handset brought to the ear (earable ability), it is the object position and its orientation which take a particular value while the actor's head coordinates are - generally speaking - not modified.

Fig. 9. Picking Up the Receiver

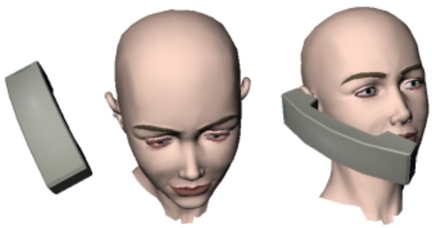


Fig. 10. Ex. of the earable ability of a phone

The system automatically defines the local coordinate system for various effectors depending on the actor morphology. Figure 11 shows the position and orientation of these coordinate systems for the "mouthable", "eyeable", and "noseable" abilities.

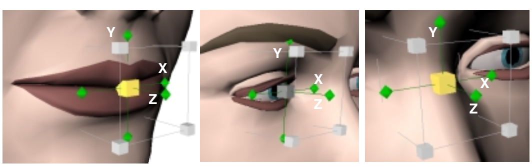


Fig. 11. Different Coordinate Systems

4 Conclusion and Future Work

In this paper, we have presented a small part of VISIONS, an innovative system for the rapid and visual authoring and prototyping of stories. In our future work, we intend to give more autonomy to the virtual characters and to integrate physical simulation, not only for enhancing the realism of the simulation, but also for simplifying even more the 3D interaction. We plan also to add non-linear scenario editing functionalities to allow the authoring or the prototyping of new forms of stories such as video-games, multimedia or interactive TV.

More information about the project is available at http://www.visions4d.com

Acknowledgement. The authors thank the European Commission for granting and supporting this RTD project in the frame of the Fifth Framework Programme.

References

1. Badler, N., Phillips, C., Webber, B.: Simulating Humans: Computer Graphics Animation and Control. Oxford University Press (1993)
2. Balet, O., Luga, H., Duthen, Y., Caubet R.: PROVIS : A Platform For Virtual Prototyping And Maintenance Tests. Proceedings of IEEE Computer Animation'97 (1997)
3. Balet, O.: An Interactive System for Virtual Cooperative Prototyping. 5th International Workshop on Simulation for European Space Programmes (1998).
4. Blumberg, B., Galyean, T.: Multi-level Direction of Autonomous Creatures for Real-Time Virtual Environments. SIGGRAPH'95 Proceedings (1995) pp. 47-54
5. http://www.c3d.com
6. Conner, D.B., Snibbe, S.S., Herndon, K.P., Robbins, D.C., Zeleznik, R.C., Van Dam, A.: Three-Dimensional Widgets. SIGGRAPH Symp. on Inter. 3D Graphics (1992) 183-188
7. Etemad, K., Chellappa, R.: Discriminant Analysis for Recognition of Human Face Images, Jl. Optical Society of America Vol. 14 (1997) 1724-1733
8. Gleicher, M.: Retargeting Motion to New Characters. SIGGRAPH'98 Proceedings (1998)
9. Moccozet, L., Huang, Z., Magnenat-Thalmann, N., Thalmann, D.: Virtual hand interactions with 3D world. Proceedings of MMM 97, World Scientific (1997) 307-322
10. Morawetz, C. L., Calvert, T. W.: Goal-Directed Human Animation of Multiple Movements. In Graphics Interface (1990) 60-67.
11. Perlin, K. and Goldberg, A.: The Improv System. Technical Report NYU Department of Computer Science (1996)
12. Polichroniadis, T.: High Level Control of Virtual Actors. PhD Thesis, University of Cambridge (2001)
13. Sederberg, T.W., Parry, S.R.: Free-form deformation of solid geometric primitives. Computer Graphics, (1986) 20(4):151-160
14. Strauss, P.S., Carey, S.: An Object-Oriented 3D Graphics Toolkit. SIGGRAPH'92 Proceedings (1992) 341-349
15. Torguet, P., Balet, O., Gobbetti, E., Jessel, J.P., Duchon, J., Bouvier, E.: The CAVALCADE System. Int. Journal. of Design and Innovation Research Vol. 2 n.1 (2000)
16. Wray, M., Belrose, V.: Avatars in LivingSpace. VRML'99, Fourth Symposium on the Virtual Reality Modeling Language, Paderborn, February (1999) 13-19

Programming Agent with Purposes: Application to Autonomous Shooting in Virtual Environment

Pierre-Alexandre Favier, Pierre De Loor, and Jacques Tisseau

Laboratoire d'Informatique Industrielle,
Ecole Nationale d'Ingénieurs de Brest,
Technopôle Brest Iroise,
CP 15 ; 29608 Brest Cedex, France,
[pfavier, deloor]@enib.fr,
http://www.enib.fr/LI2

Abstract. This paper proposes an architecture for defining and executing agents' behaviour from purposes. This architecture is used for the definition of an autonomous camera which makes automatic shooting of a virtual reality scene in real time. The user or others agents program the camera in a declarative and qualitative way. Multiples purposes can be specified. In case of contradictory purposes, the camera finds a compromise or, if not possible, leave some purposes. Multiple agents programmed by purposes generate complex and credible animations.

1 Introduction

In Virtual Reality (VR), the generation of interactive animation implying complex and dynamic scenario is a critical task [11]. The main problem is the description of the concurrent execution of various and dynamic elements. This is tackled by multi-agents approaches [4]: every autonomous agent - which becomes a virtual actor in VR [10] - has its own behavior. Agents'behaviors are executed concurrently. Interaction between agents is performed only by perception (environmental interaction) or by asynchronous communication. The more an agent will be intelligent and autonomous, the more the global scenario will become unpredictable and complex.

2 Programming Agents' Behaviour

The definition of models or architectures makes the specification of agent's behavior easier and more flexible. Furthermore, such models can get to declarative programming. In this case, the user is focused on the meaning of the behaviour rather than on implementation's details. In [3], [10], agents execute a predefined script and then have a little autonomy. The difficulty is to obtain a credible autonomous behaviour, i.e. the agent can decide by himself what it has to do at each time. Two kind of approaches lead to agents' behavior specification [9], [2]:

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 40–43, 2001.

- Symbolic approaches which come from classical artificial intelligence [7]. They are based on the assumption that the agent's perception and reasoning may be described by symbols and rules. BDI architectures [13] define such mechanisms in term of believes, desires and intentions. Such approaches are highly declaratives and allow the definition of inferences rules which manage the execution of agent's actions. The main difficulties are 1) the definition of symbols according to the agent's perception, 2) the size of the resulting database and 3) the complexity and the slowness of the inference mechanisms.
- Reactive approaches which link the sensors values to the actuators values by functions without memory [8], [6]. The reaction of an agent to a modification of its environment is then fast. The main problems are: 1) the low-level of description which generally prevents to obtain a generic and declarative specification. 2) The absence of symbol which prevents inferences and considerations about the agent's history. Then, the behaviour remains rather reactive than *intelligent* or complex.

3 Programming with Purposes

We propose a generic and qualitative approache based on a multi-agents system allowing the definition of procedures dedicated to a specific domain. The resulting agents are reactives but the behavior's specification is declarative. The characteristics elements of this approache are *purposes, trends and actions* (see fig 1):

- Purposes: purposes are declarative expressions addressed to an agent. A strength is associated to each purpose. Purposes can be multiple and contradictory. The agent *wants* to do its best according to its purposes and its perception of the environment.
- Trends: trends are informations deduced from the measurement of the purpose's satisfaction (evaluated from the agent's perception). A trend shows *what is desirable* in order to improve the satisfaction of a purpose. Multiple purposes may lead to contradictory trends. In this case the agent makes choices from the strength level and a dedicated function which relates affinity between each trend.
- Actions: each agent has its own actions, which are implemented by imperative methods. Those methods act on the environment, like actuators (walk, lie, wait, speed up, etc). A list of trends is associated to each action. This list is a post-condition stipulating *effects that can be achieved* by this action. Computing the set of purposes'trends and available post-conditions, the agent selects an action to perform.

Interpretation mechanism is running concurrently to the execution of actions (the agent *thinks* while it acts). This architecture is implemented with the ARéVi platform [12] which offers an interpreted active object language (oRis)

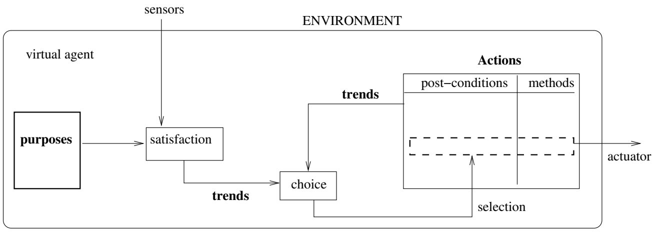


Fig. 1. Structure of an agent under purposes

[5], 3D functionalities and perception and communication mechanisms. This implementation offers generic classes, among them: *purposes, trends* and *actions*. An expert can instantiate those classes according to his domain. He must clarify the list of purposes and trends and the post-conditions close to the actions. He must also write satisfaction functions which link the agent's perception and each purpose to some trends. Next, users can specify easily the behavior of virtual agents with those purposes. Because the behavior is qualitatively specified, and because of the concurrent evolution of agents, the resulting animations are various and unpredictable. However, they remain credible according to the quality of the expert specification.

4 Autonomous Shooting

We have applied this architecture to the specification of an autonomous camera (see fig 2). This work is based on [1] which proposes a language for specifying the camera's movement. In our case, the camera finds its movement itself according to purposes which are film-making shot (*bust shot on ..., american shot on ..*). The arguments of those purposes are any graphical entities of the scene. Each entity moves freely and is perceived by the camera. Trends are based on the rendering image edge (*more on left, higher, ...*), and actions are camera movements (*dolly in, dolly out, truck, ...*). During the simulation we can program in-line multiple purposes and see the resulting camera's actions and the resulting filmed image. When possible, the camera will find a compromise between purposes, or else will choose a less critical set of actions. We can also defining new actions (with post-conditions) in line, to see the dynamic adaptation of the agent to achieve its purposes.

5 Perspectives

The autonomous shooting application is an example of our architecture. Perspectives of this work are multiples: elaboration of purposes for various domain, definition of a generic language for trends' description, formal validation or use of learning methods for automatically defining the post-conditions and the satisfaction function of a purpose.

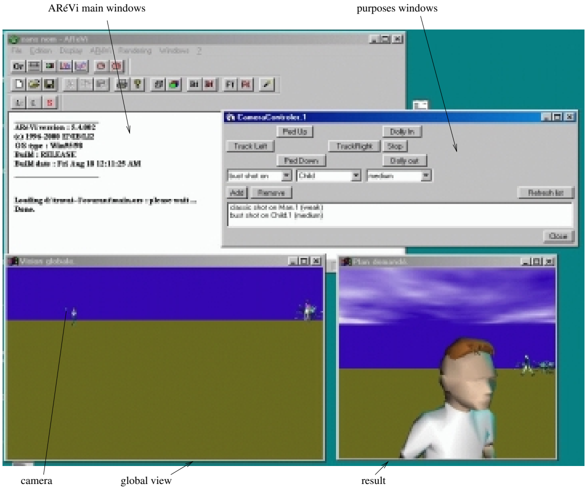


Fig. 2. Screen shot of the application

References

1. David B. Christianson et al. Declarative Camera Control for Automatic Cinematography. *AAAI/IAAI*, (1): 148–23,1996.
2. J. Bryson. Cross-paradigm analysis of autonomous agent architecture. *Journal of Experimental and Theoretical Artificial Intelligence, 12*, pages 165–189, 2000.
3. S. Donikian. Towards scenario for semi-autonomous entities. *International Conference on Visual Computing*, 1999.
4. J. Ferber. Multi-agent systems - an introduction to distributed artificial intelligence. *AddisonWesley*, 1998.
5. F. Harrouet. oris : s'immerger par le langage pour le prototypage d'univers virtuels à base d'entités autonomes, thèse de doctorat. *UBO*, 2000.
6. B. Hayes-Roth. An architecture for adaptative intelligent systems. *Artificial Intelligence : Special Issue on Agents and Interactivity*, 1995.
7. M. Ingrand, M. Georgeff, and A. Rao. An architecture for real-time reasoning and system control. *IEEE Expert*, pages 34–44, 1992.
8. P. Maes. The dynamics of action selection. In *Proceedings of IJCAI-89*, volume 2, pages 991–998, 1989.
9. M. Mateas. An oz-centric review of interactive drama and believable agents. *Technical report, School of Computer Science, Carnegie Mellon University*, 1997.
10. K. Perlin and A. Goldberg. Improv : A system for scripting interactive actors in virtual worlds. *Proceedings of SIGGRAPH'96*, pages 205–216, 1996.
11. P. Chevaillier R. Querrec and P. Reignier. Humans and autonomous agents interactions in a virtual environment for fire fighting training. *Proc. of Laval Virtual*, 2001.
12. P. Reignier, F. Harrouet, S. Morvan, J. Tisseau, and T. Duval. Arévi : a virtual reality multiagent platform. *Proceedings VW'98*, pages 229–240, 1998.
13. Michael Wooldridge. *Reasoning about Rational Agents.* The MIT Press: Cambridge, MA, USA, 2000.

Interactive Immersive Transfiction

Alok Nandi and Xavier Marichal

alterface
c/o Laboratoire de Télécommunications et Télédétection
Université catholique de Louvain
Bâtiment Stévin - 2 place du Levant
B-1348 Louvain-la-Neuve Belgium
{nandi,marichal}@transfiction.net

Abstract. The word "transfiction" has been coined to refer to an interactive narrative system where users can naturally interact with narrative machines (devices with computing power and containing databases of meaningful information). By entering the space of some camera, users are extracted out of their context: through image analysis, the visual representation of people is automatically extracted and then integrated within a pre-existing story in order to deliver in real-time scenes that mix synthetic and natural images. Such scenes are displayed on large screens placed in front of the user where she/he sees herself/himself like in a virtual mirror. She/he can then visually communicate/interact with entertaining/cultural content.

1 Introduction

The present system of "transfiction" [5] aims at extracting users out of reality when they enter the space of some camera. The captured image is analyzed, the visual representation of people is automatically extracted and then integrated within a pre-existing story in order to construct mixed reality scenes. The users' attitudes and behaviors influence the narrative (interaction layer), with the explicit intent of making the immersion (of the user's image into the visual scene) a rich experience for all users.

Contrary to many approaches to virtuality or mixed reality, the designed system does not need any dedicated hardware, nor for computation nor for the tracking of real objects/persons. It runs on standard Pentium PCs and cameras are the only used sensors. This vision-based interface approach allows complete freedom to the user, not tied to hardware devices such as helmets and gloves anymore.

Various research projects have already adopted such a user-centric approach towards mixed reality. It ranges from the only animation/command of purely virtual worlds, as in the KidsRoom [1], to more mixed worlds where users see a virtually reproduced part of themselves as in N.I.C.E. [2], and goes to the inclusion of the user image within the virtual space in order to fully exploit the potential of mixed reality. In ALIVE [3], "Artificial Life Interactive Video

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 44–47, 2001.

Environment", wireless full-body interaction between a human participant and a rich graphical world inhabited by autonomous agents is used. The Photo-realistic Interactive Virtual Environment of the ATR laboratory [4] is the system which offers more similarities with the present one since users are also reproduced within graphical environments. Where the ATR system has the advantage of considering 3D images that improve the quality of the segmentation, it remains a pure immersion system without the notion of interactive scenarios.

2 Conceptual Exploration

Mixed reality (as defined by Milgram [6]) is emerging as a major area of scientific and artistic development in today's research agenda. In this context, one of the main objectives of multimedia authoring teams is to design "interactive experiences". These are often built with image-intensive compelling applications that aim to fully exploit the potential of physical spaces and the Internet network.

However, one can claim that "Mixed Reality" is a new term for an old praxis illustrated in literature and philosophy for centuries. Plato in the West will be the supporting example, but one could look at writings in the Oriental civilisations (India, Tibet, China, Japan...). Different reality levels (physical, cognitive, psychological) are layered into mixed and complex reference systems. This evolving framework plays a role in definitions of concepts (immersion, intuitive, spatiality ...) which are reviewed in today's technological context.

In "*Respublica*" (which means the "public thing", one could relate to the concept of audience experiencing a "common" public object), Plato describes the classic scene in the cave. People are sitting close to a fire and their shadows are projected on the wall (low technology and image projection are used here and can already be seen as a source of confusion between reality and representation). The "spectators" are fascinated by these moving images. They go out of the cavern and the light of the sun makes them in a way blind so they go back into the cave. One could argue that they become twice victims of the "light", inside and outside the cave: their perceptual system has to evolve to realize that the projected images, the shadows, are related to the movements of their bodies.

This case study of Plato's cave serves one purpose: to bring the notion of "intuitiveness" into perspective, outlining its definition in relation to contextual evolution and cognitive maturity. Humans behave intuitively in spaces, and the nature of space is changing because of its materiality. Space is getting more into virtuality, through its representation (3D, computer imaging...) as well as through its experience (heterogeneous network, relation to distant places, time difference contraction...). Confrontation with images, with "transformed copies" of real people (the shadows) led Plato to explore concepts of representation and interpretation, in an allegoric way.

Recent technological developments allow one to explore the notion of "transfiction" introduced in [5]. Visual screens are the output zones of sequences of images and narrative instances. But these can also be input zones for interaction in artificial spaces. An "interactor" gets immersed and uses his/her body "as a joy- stick" to trigger off a set of narrative sequences and to interact with the artificial spaces. The concept of immersion moves from a cognitive (cinema offers the ideal example) to a physical experience. The psychology of the projection of a viewer into the character of the film and the interpretation of his/her role is now to be reconsidered in the framework of Mixed Reality. The processes of perception and cognition are located in both real and virtual space at the same time. The interaction with an alter ego into a "magic mirror" (image of one's self projected onto the screen) allows for intuition to lead the activities in a mixed reality system. The "transfiction" system allows users to interact with the virtual space through gestures and movements. Speech can offer another modality for interaction and will be analysed in some future work.

In these hybrid action spaces, narrative modalities are investigated through tracking, position detection and image recognition systems. The technical architecture described hereafter allows the implementation of "narrative graphs" and opens the possibility to analyse the cognitive and visual actions and reactions to an immersive intuitive interactive system.

3 Technical Architecture

In order to provide users with such an interactive immersive experience, the technological challenge is to gather all needed subcomponents and issue a real-time implementation of the system. To compose all the visual objects (the "real" and computer-generated ones) within the final mixed reality scene and to allow for interactivity, the MPEG-4 standard [7] can be used as the transmission layer.

In addition to the composition and transmission facilities, the system also achieves real time segmentation of moving objects captured by cameras and automatically extraction of descriptors (MPEG-7 [8] like) that are used to describe the behavior of visual objects and interact with the scenario. Thanks to a client-server architecture based on the Internet Protocol, the system is very flexible and allows any screen to access any resource it needs. A phenomenon of ubiquity is therefore provided since two or more screens may simultaneously access the same camera.

4 Narrative Graphs

Any interactive narrative is established as a (narrative) graph. Every node of the graph provides one with a scene (composed of various real/virtual objects) along with a list of events to be trigged. For instance, the system may be told to look for a person to 'touch' a particular graphical object which is a door, or to detect two persons moving 'fast' in opposite directions, or simply to wait for 15 seconds. According to the detected trigger, an action occurs. The action is a move to some next node of the graph, where another scene is depicted and other triggers are searched for. The scene can be a completely new one or the same one from the previous node with just some additional (or suppressed) graphical element (scene refresh or update).

It is crucial to note that the evolution of the narrative is different for every screen of the system, i.e. for any player or for any set of interactors in front of the same large screen. Narrative graphs are thus dealt with in an autonomous way in order to allow different users to enjoy the same story at different moments and with different interactions.

5 Conclusion

Transfiction was initially coined for "transportation into fictional spaces". An important driving factor of the "transfiction" concept is to look for immersion solutions where the interactions are natural in order to provide the user with a rich "interactive experience". The provided demonstration aimed at publicly assessing the interaction modalities while showing that the quality and the richness of the experience do not necessarily require high end computing.

Acknowledgment. The present research is essentially achieved within the art.live project. This project (*http://www.tele.ucl.ac.be/projects/art.live*) is part of the 5th framework programme of the European Commission: Information Society Technologies (IST, project 10942).

References

1. Bobick, A., Intille, S., Davis, J., Baird, F., Pinhanez, C., Campbell, L., Ivanov, Y., Schütte, A., Wilson, A.: The KidsRoom: A Perceptually-Based Interactive and Immersive Story Environment. PRESENCE: Teleoperators and Virtual Environments, 8(4), August 1999, pp. 367-391.
2. Roussos, M., Johnson, A., Leigh, J., Vasilakis, C., Barnes, C., Moher, T.: NICE: Combining Constructionism, Narrative, and Collaboration in a Virtual Learning Environment. Computer Graphics, 31(3), August 1997, pp. 62-63.
3. Maes, P., Darrell, T., Blumberg, B., Pentland, A.: The ALIVE system: Wireless, full-body interaction with autonomous agents. Multimedia Systems, 5(2), March 1997, pp. 105-112.
4. Kim, N., Woo, W., Tadenuma, M.: Photo-realistic Interactive Virtual Environment Generation Using Multiview Cameras. in Proc. SPIE PW-EI-VCIP'01, vol. 4310, Jan. 2001.
5. Nandi, A., Marichal, X.: Transfiction. Virtual Reality International Conferences 2000, Laval, 18-21 May 2000.
6. Milgram, P., Colquhoun, H.: A Taxonomy of Real and Virtual World Display Integration. in Mixed Reality, Merging Real and Virtual Worlds, Y. Ohta and H. Tamura editors, Ohmsha-Springer, 1999.
7. MPEG Video group: MPEG-4 Video Verification Model, version 11.0. ISO/IEC JTC1/SC29/WG11 (N2178), Tokyo, March 1998.
8. MPEG group: MPEG-7 description definition language document. version 2.0, ISO/IEC JTC1/SC29/WG11 (N2997), Melbourne, October 1999.

Interactive Storytelling: People, Stories, and Games

Andrew Glassner

Writer
andrew_glassner@yahoo.com
http://www.glassner.com

Abstract. Storytelling, people, computers, and digital communications are becoming increasingly interwoven. The idea of using procedural techniques to involve people in stories is enormously attractive, yet actually finding a way to create interactive fiction that achieves both artistic and commercial success remains elusive.
In this short paper I will briefly discuss a few of the relevant issues for designing interactive fiction. I discuss the need for story structure, and the difficulties of asking people not trained in acting to become improvisational actors. I then present an idea called the *story contract* that describes some important traits of successful fictive experiences. Finally, I discuss some of the inherently contradictory needs of stories and games. [1]

1 Introduction

Stories are a vital part of human culture. We use stories to entertain each other, to pass on our cultural and personal histories and values, and to make sense of our world. Stories give us continuity and context, and help us to understand not only what the world is like, but to discover our own places in the world and how to live our lives.

Driving head-first into this millennia-old human tradition are the radical new fields of computing and electronic communications. Just as these technologies have changed commerce and industry, they are already changing our social lives.

What will happen to stories? Already, many people are trying various experiments to bring new technologies to traditional narrative, but so far most have failed either artistically or commercially.

It seems clear that there's a lot of money to be made by merging stories with technology. People spend significant time and money on stories through their enjoyment of films, television, books, theater, and other sources of fiction. The major American movie studios made about $7.7 billion in 2000. Contrast this to the approximately $6 billion taken in by the manufacturers of electronic and video games, and it's clear that the new technologies have caught on fast. Many

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 51–60, 2001.

analysts expect the games industry to overtake the film industry in net revenue within a few years.

The common dependence of both modern media and gaming on computer technology is drawing these two fields together rapidly. When they collide, a dramatically new form of popular entertainment will emerge: *interactive fiction*.

Stories are part of what makes us human. Great storytelling is popular art. It is also commercial: we crave stories in movie theaters, in books, on television, in newspapers, during business meetings, and almost every facet of our lives. On the other hand, games are a popular and enjoyable way to share in the company of other people. We play games with each other, and follow professional players and teams as they compete. Stories and games have rarely been combined successfully. This is about to change.

As stories and games merge in the medium of computers and world-wide communication, we will see the birth of the hybrid interactive story. This will be a mature form, ultimately capable of the same depth of expression as the greatest literature or film, but also capable of the same visceral, immediate gratification of the most challenging sports and games.

This new form has enormous commercial and artistic potential all over the world. Many authors, game designers, and companies have started to move in the direction of merging games and stories.

Unfortunately, nobody has yet made a commercially successful interactive story worthy of the term. Why?

It turns out that games and stories are in many ways incompatible.

There are many reasons for this. For example, we typically enjoy stories passively, and usually as individuals, even while sitting in a group audience. On the other hand, we participate in games actively, frequently in a social environment. Many other aspects of stories and games are directly contradictory. Because they are both mature fields, some of these contradictions are very subtle, yet go right to the core of what makes each form work.

2 Stories

What is a story? At its core, a traditional story describes a sympathetic character who want something desperately, and confronts a series of escalating obstacles to obtain it. The character may in fact not reach his or her goal in the end; it's the struggle and not the conclusion that forms the heart of a story.

Aristotle identified three basic forms of conflict: man vs. himself, man vs. man, and man vs. the world. The first form has given rise to many great dramas, and the concept that many great men (and women) carry within them the seeds of their own destruction. The third form is seen often in stories that pit a person against the elements, often in the form of violent nature (e.g. storms at sea, incoming asteroids, floods, volcanoes, etc.).

The second form is typical of most commercial drama today: a hero (or protagonist) faces off against a villain (or antagonist). There are lots of story forms that can be built on this foundation. For example, the hero and villain might both want (or need) the same thing, and race each other to obtain it. For example, an atomic bomb is lost when a sub goes down, and the good guy and the bad guy are both eager to retrieve it. In another common realization, the villain takes something from the hero. For example, the villain could kidnap the hero's spouse. The villain doesn't actually want the stolen goods per se, but only to use them as a weapon against the hero. The conflicts can be much more subtle and personal, of course.

It is well known that many stories in the Western tradition share a three-act structure. In essence, we meet the characters and discover the hero's problem, we see the hero deal with increasing obstacles to resolve the problem, and eventually there is a climax where the hero risks everything to achieve his or her goal. A richer structure was articulated by Joseph Campbell in his book, *The Hero's Journey*.

These don't exhaust the possibilities, but it is clear that good fiction has good *structure*. Just as a building stands up over time only if there is a solid structure inside, so it is with fiction.

2.1 The Story Contract

Although there are many different fictional forms, they all share a few things in common. I characterize three of these most important elements as *the story contract*. There are two responsibilities for the author, and one for the audience.

The first clause of the story contract states that *the author is responsible for the psychological integrity of the main characters*.

Interesting characters are those that we can understand and empathize with. This requires us to be able to relate to them as people. Of course, people are complex, filled with contradictions, and display unexpected behavior from time to time. But human behavior is reliable enough that we are able to form relationships; if we couldn't trust people deeply we wouldn't be able to form families and sustanined, loving relationships.

When people begin to behave erratically we are become concerned for them, and sometimes encourage them to find the physical or psychological cause for their behavior. When people start to act very unpredictably, we often avoid them, and definitely avoid entanglements.

Although the hero of a story needn't be a potential friend (indeed, some stories feature anti-heros that are decidedly unlikable), successful heros are fascinating. They are complex and interesting people, and we get drawn into their heads because we see enough of ourselves that we can relate and care.

If we're not inside a hero's head as the piece progresses, then the story can fail in many ways. First, the audience can simply not care, which is a disaster for any story. Second, the audience can tune out (either literally or mentally), which is just as bad. If the audience doesn't care about the hero, then when the final conflict comes the whole emotional pressure of the story will fall flat, because the audience isn't personally involved in the trauma faced by the hero.

Writers create their heros with great attention and craft so that audiences will empathize with them. A hero's actions may be surprising, but there is an inherent consistency that eventually reveals itself. If the audience is manipulating the hero, or making choices for the hero, then that consistency is very easily lost.

In so-called *branching narratives* the audience is often asked to make decisions for the main character at moments of maximum stress and conflict. This is *exactly* the time that the writer's control is most essential! How characters behave under stress is what reveals their real personality. If the audience chooses one way at one time, and another way at another time, then it's unlikely that the main character is going to be perceived as having any kind of stable personality.

If an audience is going to control a character, the mechanism needs to be much more carefully planned than the blunt instrument of branching narrative.

All that I've said above for the hero is also true of the villain, if there is one. In fact, it may be the case that the psychological integrity of the villain is even more important than that of the hero, since a villain who acts in truly erratic ways doesn't gain an audience's respect or sympathy, and therefore never receives the kind of negative emotional involvement that gets released at the end in the cathartic climax. An erratic villain is like a hurricane; it's certainly a problem, but you can't really hate a hurricane or wish it ill; you just want it to stop or go away.

The second clause of the story contract states that *the author is responsible for the sequencing and timing of major plot events.*

This is just an encapsulation of common sense. If we're in a kidnapping story, then it's important that the victim be kidnapped before the police go searching for the victim. The purpose of this clause is to make sure that cause precedes effect.

It's important to state this explicitly because there have been a number of computer games and story forms that give the player an opportunity to manipulate the world in ways that contradict this principle. In fact, some games that otherwise maintain a consistent world state sometimes dramatically and unexpectedly modify that state in order to reset the world in anticipation of a player's next action. Hypertext in particular suffers from this problem, since it's very hard to give the reader a sense of control and exploration without giving up some control over the linear sequencing of the narrative.

The second clause basically says that we shouldn't call the fire department unless there's a fire.

The third clause of the story contract states that *the audience must allow itself to be emotionally moved.*

This is required because in order for a piece of art to work, it needs to be able to speak to us on an emotional and personal level. If we remain at arm's length, or emotionally disconnected, then the piece cannot reach us. Sometimes a audience member will criticize a piece of art by saying that it didn't "move" him or her. Of course, a work can only move us if we open ourselves up and allow it to.

This is an unusual demand to make of an audience member. Most of us spend our lives with our emotional shields up to protect us from the random people and events of the world. We typically only allow access to our core beings to people we know and trust. That is why a loved one can cause so much more pain with a casual remark than an stranger on the street who hurls a blunt insult.

But when we attend a piece of art (by reading a book, attending a play or concert, going to a film, or ultimately playing a video game), we must open ourselves up emotionally to allow the piece a chance to move us. A good piece of art can reach us on many levels: emotional, spiritual, intellectual, and more. To give it that opportunity, we must allow the work to manipulate our emotions.

Of course, the audience member who does this is ultimately in control, because he or she can retract his or her consent and restore the emotional barriers at any time. Typically when people get too afraid in a horror film they quickly distance themselves from the scary story and become very interested in the technology of the special effects, or wonder how the fake blood is made and stored. These are natural defense mechanisms that we use when we find that we have opened ourselves up beyond a comfortable point, and we don't find that the rewards of staying open exceed the discomfort.

The value of the story contract is that it gives us another tool for examining new types of fiction and evaluating their possible success.

Notice that despite many experiments and attempts, the two most popular forms of non-linear storytelling today (branching narrative and hypertext), have failed to achieve mainstream success. One reason for their commercial failure is that they both break at least one clause of the story contract. Branching narratives virtually all invite the audience to participate by selecting actions for main characters at important moments. Thus they break the rule that the author must control the psychology of the main characters. Hypertext breaks the rule of causality, because hypertext inherently allows the reader to explore the story in an unpredictable order.

Although these are fun novelty forms, and can be used to create serious art, experience has shown that these are not good structures for creating mainstream stories.

3 Home Acting

It seems clear that some kind of audience participation is an essential part of anything that we might call interactive fiction. If we can't have the audience making decisions for the main characters, then how about instead putting them directly in that role? There are a variety of commercial games in the market now that essentially place the player in the role of the lead character in a drama, either for a solitaire game alone with the computer, or in the midst of a larger group of people.

If the person is simply play-acting, then this can be fun within limits. If the idea is to actually have the person help create the drama, it usually fails.

The essential problem is that these games are asking their players to be improvisational actors. The problem with this approach is that improv is a skill, like playing basketball or navigating a sailboat. Improv is an acting technique that, when performed on stage, is actually managed within a context of rules and conventions. Performers study and practice those rules and conventions until they're second nature, just as a musician practices on his or her instrument until the mechanics of playing can be transcended during a performance. Furthermore, improv doesn't admit lengthy pieces: most improv scenes last only a few minutes, because it's difficult to make the forms work much longer than that. Simply put, improvising well is a learned skill. Most people do not know how to ride a racehorse, play the saxophone, or improvise comedy or drama.

So let's take improv out of the equation. To remove the difficulty of improvising characters, scenes, relationships, and the rest, let's build on the fruits of someone who's done it already.

Many great plays are widely available. Shakespeare's *King Lear* is undoubtedly a great play. So if we believe that people enjoy being the stars of fictive experiences, why don't people put on *King Lear* at home, just for their own enjoyment? It wouldn't be too hard to sew some costumes and build a simple set. But we could even do away with all of that, and just gather a group around a table and read the play out loud. Such reading groups exist, but not many. Few people participate in this activity. Why not? It has everything: great characters, great lines, a great plot. It's all there and ready, all we have to do is say the words.

The problem is that this form of acting, just like improv acting, is hard. If you ask a child if he or she is an artist, odds are they will say yes. If you ask an adult, odds are he or she will say no. What happened? As we get older, we recognize that those who study a craft or art develop certain skills that allow them to execute their work with grace and control. The average adult who does not play the trumpet will recognize that a great trumpet player has skills that he or she does not have, and moreover, does not want to put in the time and effort to acquire. After all, we can only master so many things in life. If we choose to become a master chef, we may not be able to also become a master mountaineer, a master sculptor, a master politician, and so on.

Acting is a complex skill. And it's a peculiar one, in that it requires the actor to experience emotions. Of course, many emotions are unpleasant. And that leads us to the problem: most people do not want to deliberately experience unpleasant emotions as a form of recreation! Consider playing Lear himself in *King Lear*; you're slowly going crazy, your daughters are betraying you, and your life is falling apart. Very few people looking for an evening's fun would choose to place themselves in this situation! Lear is of course a fascinating character, and watching a good performance of the play can be a riveting and powerful experience. But *performing* in the play is not something that most people would find to be particularly fun.

The lack of skill is not the major problem here: many people enjoy playing sports that they're not particularly good at, as long as the companionship is good and the activity itself is enjoyable. It's that acting requires feeling unpleasant and undesirable emotions, which people are naturally reluctant to assume and display unnecessarily.

So we have two good reasons not to put untrained actors into a lead acting role. First, if the role requires improv, then most people will find that they don't have the skills to make it work, and those who indeed are practiced at improv can't keep it going for more than a couple of minutes. Second, even when the entire piece has been authored and written down, people don't want to perform it because it requires them to be skilled actors, and because they have to deliberately take on and present negative emotions.

Notice that home acting doesn't require any technology. Ever since plays were first printed and available on paper, people anywhere could have performed them for each other as a recreational activity. The fact that almost nobody does it is an important lesson for designers of interactive fiction.

In summary, most people are not skilled actors, and do not enjoy it, regardless of whether they are making it up or working from a masterpiece.

4 Stories, Games, Puzzles, and Toys

I like to break down the field of interactive entertainments into four categories.

Stories are narratives that involve plot and character.
Games involve the development of skills and usually involve competition.
Puzzles are games with a predetermined solution.
Toys have no fixed purpose.

Some toys are tools in other contexts. For example, many people make their lives working on boats. But a sailboat used for recreation is simply a toy; it's there to be used at whim, simply for pleasure. Children often use plastic construction tools, kitchen equipment, and other toys that simulate the tools used by adults.

These different forms have very different needs. To suggest the scope of these differences, I'll focus below on just three of the conflicting needs of stories and games. I've chosen these two in particular because almost every form of interactive fiction with audience participation that has achieved any measure of mainstream success has attempted to blend stories with games. As we'll see, this is a difficult proposition because the needs of the two topics are so different.

4.1 Communicating

Certainly great stories can be incredibly entertaining. From the *Iliad* to the to the *Bhagavad-gita* to *Citizen Kane*, a great story grabs us by the gut and doesn't let go until it's through. But stories are also capable of teaching, and all three of the stories listed above share fascinating information about human nature, philosophy, and the world around us along with their entertainment.

Through stories we learn about the lives of people in other places, times, and walks of life. We can read a factual description of life in the desert during World War I, but seeing *Lawrence of Arabia* (or reading *Seven Pillars of Wisdom*, the book on which it was based), for all its invention and inaccuracies, gives the subject immediacy and vibrancy. Most of us will never know what it was like to live in Hellenic Greece, or live as a pop music star in 2001. But stories can help us see what those people faced (or currently face) in their everyday lives, and expose us to different ways of living in the world.

Games, on the other hand, are primarily about the experience of the moment, and challenging of the self. The communication is from the master to the student in the form of teaching skills and technique. It's up to the student to internalize the teaching and develop his or her skills. Most games are about remaining focused in the moment, and acting skillfully.

Of course, both stories and games are entertaining.

	Stories	Games
Communication:	Learning	Experience
	Vicarious living	Skill mastery
	Diversion	Diversion

4.2 Actions

What is running through our heads when we're reading a book, or watching a character on the stage or screen? In a great story, we're immersed in the moment, living out the person's life with them. But rarely are we totally lost; more usually we are in the position of a privileged insider. We may even have special information that the hero doesn't know; for example that the villain is waiting just behind the door that the hero is about to open.

In other words, we're aware of what's happening and the larger context even while it's going on. In a film or play, we are pulled along in real time by the driving narrative. In a book, we can set down the text any time and ponder what's happening, and even discuss it with other people.

When there's time, we consider the situation from the point of view of the character with whom we are empathizing. We think about the costs and rewards of different actions, and we weigh the consequences. This lets the magnitude of the risks taken by the character work into our consciousness, letting us bind with the hero ever more deeply.

Even in real-time forms such as film, we often feel increasing tension as a character builds up to taking a huge and risky action with scary consequences; it's our knowledge of what's likely to come that fuels our anticipation.

In contrast, sports-style games offer little to no time for thinking. They're all about immediate and optimistic action. When skiing, for example, you may come across a patch of ice. Your action in response must be decided upon and executed in that moment; there's no time for deliberation.

The choices made during this kind of real-time sporting game are impulsive and subconscious; they are affected as much by our emotions than our intellect. We make a decision based on experience and then execute on that choice as skillfully as possible. It is this immediate, thrill-of-the-rush feeling that makes so many high-energy sports enjoyable, from tennis to sailing.

	Stories	Games
Actions:	Thoughtful	Impulsive
	Conscious	Subconscious
	Deliberate	Spontaneous
	Intellectual	Emotional
	Weighted	Hopeful

4.3 Rules

What are the rules of the world in which a story takes place? The rules can be anything. There are two principal types of rules to think about in fiction: the physical rules of the world, and the social rules of behavior and society.

The first type of rules are most frequently changed and explored in fantasy and science fiction. In the *The Wizard Of Oz*, we start off in a world that looks and behaves in familiar ways, but then we go into a world that contains witches, flying monkeys and talking lions. The film *The Matrix* travels from an everyday world of the near future to a strange and hidden world where nothing is as it seems.

Social rules can be obvious or subtly different than our everyday world. Most of us are familiar with the fictionalized world of organized crime, where the rule of silence is lethally enforced. Often the rules are based on the characters themselves: there are certain subjects of family and sexuality that are just too dangerous to touch in *Cat On A Hot Tin Roof*. It is only by breaking those rules that the characters are able to change.

In fiction, as in real life, we often do not know what the rules are. They are fundamentally unknown to us, established by people and organizations we have never seen, and enforced by people and organizations who are not primarily concerned with our individual interests. Some rules are overt, and codified in law, but others must be discovered by trial and error.

Part of the culturization process we go through as children is to become sensitive to the boundaries of allowed behavior. These rules are often not written down anywhere, and in fact are flexible and change over time.

Some of the rules that we follow in real life, as in fiction, are internally driven. We may be on a road that allows us to drive at a certain speed, but we may drive well below that speed because we are tired or don't have confidence in our vehicle. This is an internal rule that we apply to ourselves. Many moral and ethical behaviors result from these type of rules. When it's convenient to lie, many people choose not to: it's strictly a result of an internal choice.

Some internal rules are driven by habit: we may always descend a flight of stairs right foot first, for example. Others are driven by resolve, such as when we refuse another piece of chocolate cake because we're trying to lose weight. We may choose to break these internal rules if we want, and in some circumstances nobody will know but ourselves. Sometimes we don't know what rules we live by until they're tested, and the same is true of fictional characters. One of the joys of well-written fiction is watching characters behave in stressful situations, and slowly learning more about their inner makeup.

Games have a completely different rule structure, which is far more formalized and explicit. This structure is as consistent for video and computer games as it is for sports.

The rules are laid out in advance, and are explained to all the participants. Typically the players are asked if they understand the rules, and if they do not, they are explained again. The rules are often stated clearly in printed rulebooks, distributed before the event begins.

These rules are external, in the sense that someone beyond the players has decided upon them and laid them out. In some children's games the rules are flexible and invented on the fly, but generally the rules are set by an external authority before the game begins, and all players adhere to the rules.

Compliance with the rules is not only mandatory, but there are external referees to enforce compliance and adjudicate disputes. Once the game is begun, the rules are fixed and must be obeyed.

	Stories	Games
	Unknown	Announced
	Internal	External
Rules:	Discovered	Explained
	Subjective	Objective
	Moral	Refereed
	Shifting	Fixed

5 Conclusion

In this short paper I have presented just a few of the issues that confront designers of interactive fiction. My goal was certainly not to be exhaustive, but rather to suggest the shape of some of the the problems, and present a few selected thoughts for addressing pieces of it.

In particular, I've discussed story structure, and the difficulty of asking normal people to become improvisational actors. I've presented the concept of the story contract, and argued that the failures of branching narratives and hypertext in the mainstream may be at least partially traced to their contradiction of at least one of its three principles.

I've also presented some of the characteristics of stories and games, and suggested that these elements are in fact often contradictory.

I believe that there is a bright future for some form of participatory narrative, but it will not come from simply plastering stories and games together and hoping that the result is somehow enjoyable.

Rather, high-quality, commercially successful interactive fiction will result from a principled understanding of stories and games, and a thoughtful and principled technique for blending them together while respecting what people actually find to be enjoyable and rewarding stories and activities.

An Authoring Tool for Intelligent Educational Games

Massimo Zancanaro, Alessandro Cappelletti, Claudio Signorini, and Carlo Strapparava

ITC-irst
Panté di Povo – 38050 Trento Italy
{zancana|cappelle|signori|strappa}@irst.itc.it

Abstract. In this paper[1], we describe a frame-based production rules system that works as the Artificial Intelligence Engine of an educational computer game. We discuss the need of an authoring environment clearly separated by the game in order to allow a technical staff without any skill in either AI or Computer Science to encode the "intelligence" of the game. Finally, we briefly introduce two graphical interfaces for authoring and testing frame hierarchies and production rules. The production rule systems and the authoring tools have been developed in the context of a project funded by the European Community to develop a prototypical educational computer game.

1 Introduction

Today there is a wide acceptance on the role of AI to build more compelling computer games ([1]), yet very little concern has been shown on letting content experts rather than programmers design the "intelligence" of the system. The authoring issue gains dramatically importance in the design of educational (and yet engaging!) computer games, where you would like to let content experts or an editorial technical staff to define and tests the rules of the game. Indeed, in the near future it might be valuable to hire professional script writers even for non-educational games.

In this paper, we briefly discuss our experience in the design and implementation of a rule-based engine to be used in an 3D on-line educational computer game and its authoring environment. This work is part of a project called RENAISSANCE[2] funded

[1] A shorter version of this paper has been presented at the workshop "Artificial Intelligence and Interactive Entertainment" at the AAAI Spring Symposium, Stanford, March 2001.

[2] The partners of the RENAISSANCE project (IST-1999-12163) other than ITC-irst are Giunti Multimedia, one of the biggest Italian publishing companies, as the main contractor; Blaxxun

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 61–68, 2001.

by the European Community in the action line of “access to scientific and cultural heritage”.

2 The RENAISSANCE Project

The aim of the RENAISSANCE project is to develop a computer game that makes use of high quality 3D graphics and engaging interaction yet able to deliver scientifically validated contents. The long term goal is to experiment an innovative pedagogical approach: delivering culture in an effective and amusing way at the same time.
The game is conceived as a 3D-based multi-user role-playing virtual community over the Internet.
The game environment is the renaissance court of Urbino in central Italy around the first half of the fourteenth century. The term Renaissance describes the period of European history from the early 14th to the late 16th century, the name comes from the French word for rebirth and referred to the revival of the values and artistic styles of classical antiquity during that period, especially in Italy.
This scenario was chosen because life in that period was subject to complex and subtle behavioral rules so precisely defined that have been codified in handbooks, in particular the famous “Book of Courtier” by Baldassarre Castiglione, published in 1528.
The players, as courtiers, have to increase their social positions and compete to obtain the Duke and Duchess’ favors. The ultimate goal is to enable users to experience, as realistically as possible, the complexity of the social life during that fascinating historical period while having the same fun of playing a “state of the art” video game.

The score of each player is expressed in terms of his fame, fortune, faith and force which can vary according to the his “opportunistic” behavior in different situations. The “intelligence” of the games resides in a rule-based system (called the Evaluation Engine) that computes the “effect” of the players actions in the virtual world.

In the next section, we briefly introduce the system architecture focusing on the internal structure of the Evaluation Engine. Then, in the last section, we will describe the authoring environment actually used by an editorial staff to encode the rules of life in our virtual renaissance court.

Interactive a german-based company whose main business is 3D-based virtual environments over the Internet and Iridon Interactive a Swedish company that produces and distributes computer games.

3 The Game Architecture

The RENAISSANCE game is a 3D-based multi-user role-playing game over the Internet. The 3D rendering engine is local to each client and a Virtual Community Server (VCS, for short) is in charge of maintaining the synchronization among the different clients. At each user action, the VCS computes the visible effects (in terms of rendering) and communicates the changes to the other clients. The Evaluation Engine, instead, is in charge of maintaining the coherence of the world from a semantic point of view: at each user action, it computes the "pragmatic" effects both for the user that performed the action and for the rest of the world. The Evaluation Engine is updated and queried by the VCS through a message protocol based on KQML [2].

3.1 The Evaluation Engine

The Evaluation Engine is based on a frame system called CLOS-i built on top of CLOS (the Common Lisp Object System) exploiting the meta-object capabilities of this language. In designing CLOS-i our aim was to develop a "light" knowledge representation system yet efficient enough to be used in a complex scenarios. The production rules system employs an implementation of the RETE algorithm [3] modified to be used together with a hierarchy of frames.

Rules and frames are two complementary knowledge representation schemes. There are several attempts to integrate these two approaches, but few efforts (in particular, [4] [5]) have been made to incorporate the terminological knowledge of frame-base systems into a rule-based paradigm. We think that this approach improves conventional ruled-based programming from many points of view. In particular, the pattern matching operation is based on terminological definitions, not just on symbols (like in OPS5, for example) and conflict resolution can be based on well-defined specificity relationship among rules. Moreover, this approach encourages the development of a large and coherent knowledge base that is shared among the rules.

3.2 Example of a Situation

We discuss here an example of a situation modeled in the very first KB of the RENAISSANCE game: "every day at 10am an evening dinner with the Duke is organized. Each courtier with more than 500 points of fame receives an invitation. The dinner starts at 7pm. Courtiers which have got an invitation and do not attend the dinner loose 100 points of fame". In order to model the organization of the dinner, the more general frame of **activity** has been defined so that starting and finishing of activities can be implemented as general rules. The **dinner** frame is defined as a subframe of **activity**, it has no slots because it has no special properties. Indeed, we need this new frame in order to write a more specific rule: every day at 10am the dinner (but not necessarily all the other activities) is scheduled; the rule **dinner_organization** is fired every time an instance of **set_time** is received with 10 as value of the **hour** slot; the action is the creation of a new instance of **dinner**.

The rule **dinner_invitation** is triggered by the creation of an instance of **dinner**, the other condition is that it exists a courtier with more than 500 points of fame. An action for the creation of an instance of **invitation** is built for any such courtier. The rule **invitation_notify** takes care of communicate the events.

Once the dinner starts (according to the general rule **activity_start**), the rule **dinner_attendance** will fire on each courtier for which an instance of **invitation** exists and it will decrease the his/her fame.

4 The Evaluation Engine Authoring Environment

We decided to employ a frame-based production rule system because our main concern was to allow a staff of technical editors of writing the "intelligence" of the system. Other researchers showed that productions rules are a tool powerful enough for describing human cognition (see for example, 6) and simple and intuitive enough to be understood by naïve users (see for example, 7). Yet we realized that we had to provide interactive tools to let the editors to graphically manipulate the frame-based system and interactively test the rules independently from the game engine in order to let the editorial work to proceed parallel to the work of the programmers and to the work of the designers.

We implemented two graphical interfaces: the Knowledge Base Editor and the Knowledge Base Shell.

The former allows to graphically manipulate frame hierarchies, to define and to edit frames and slots and to write rules. It exports the knowledge-bases as XML files.

Figure 1 depicts a snapshot of the KBE. The main window is divided in two parts, on the left window the user can choose whether to work on the frame hierarchy or on the set of rules; the right window is used to edit the particular frame/instance/rule selected on the left window. In the snapshot, the frame *courtier* is selected on the left window. Each frame has a number of slots that represent the attributes of the concept. A frame automatically inherits the slots of its parent frame[3].

[3] At present, multiple inheritance is not allowed. This feature can be dealt with in the present evaluation of the Evaluation Engine yet it may led to very inefficient and confuse knowledge bases.

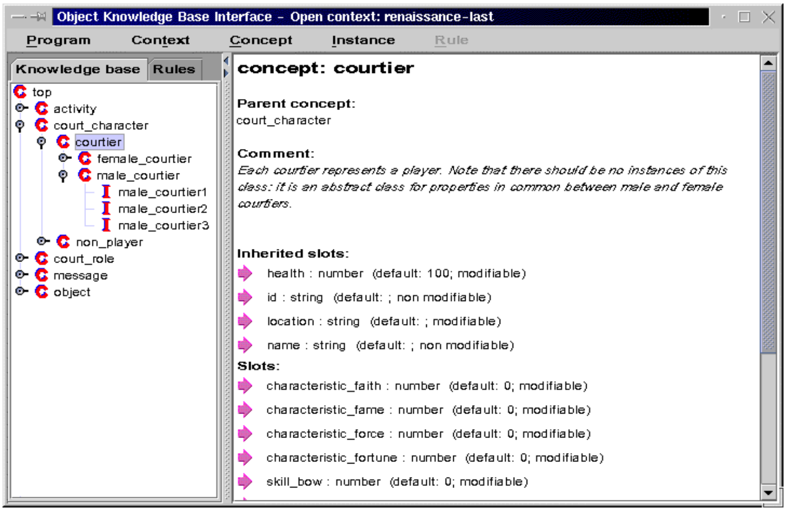


Fig. 1. A snapshot of the KBE main window.

Editing the frame hierarchy means to edit frames and slots (i.e. working on the terminological part) or to edit the instances of an already defined frame (usually, instances are created, modified and deleted at run time by the Evaluation Engine, yet it can be useful to have some pre-defined instances, for example non-player characters, furniture, etc.). These two activities can be interleaved, KBE is able to maintain the whole knowledge base consistent (for example, deleting a frame means removing all its instances; more subtly, it sometimes requires to remove a slot from another frame and in turn all the corresponding slot values from its instances). Usually, KBE performs silently these operations, yet when the amount of deletions is big it warns the users before continuing. Moreover, the interface has been designed to minimise the likelihood of having inconsistent knowledge bases. For example, the user can never create a dangling frame (that is a frame without a parent): the only way to create a new frame using the interface is to add a child frame to an existing frame[4].

KBE supports the rules writing task as well (see figure 2). The task of writing rules logically occurs after the creation of the knowledge base (because the left-hand side of a rule is expressed in terms of frames and possibly instances.) In our experience, however, the two tasks are highly interleaved: a first sketch of the frame hierarchy is necessary before any rule can ever be conceived, yet the actual writing of rules usually suggests new frames or even a different organisation of the hierarchy. Therefore, we designed the interface with the goal of making easy to the user to interleave the two tasks. In order to avoid inconsistencies as much as possible the rules are composed by direct manipulation: before using a concept in a rule, the corresponding frame has to be defined in the hierarchy. As in the task of knowledge base editing, a lot of checks

[4] The very first frame is automatically created by the system and it's name is always *top*.

are performed automatically to maintain consistency: for example, if a frame is deleted, all the rules that use the corresponding concept are deleted as well.

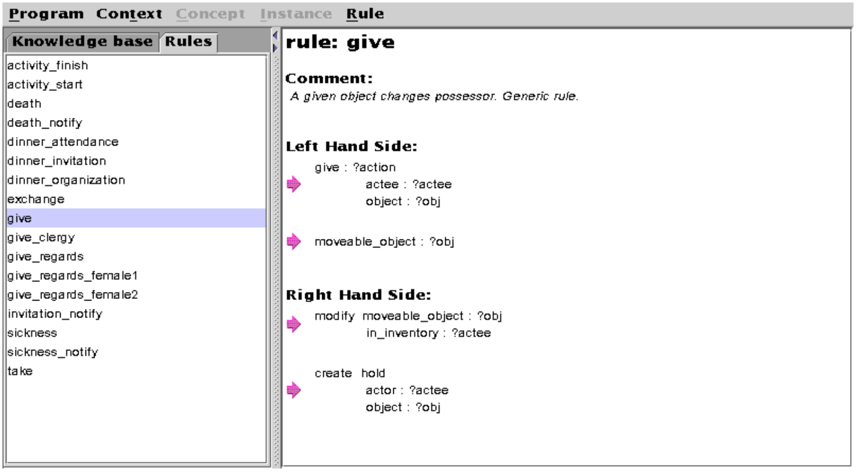


Fig. 2. Rule editing with KBE

The second tool of the authoring environment is the Knowledge Base Shell (or KBS, for short). It communicates with the Evaluation Engine in the very same way as the game will (i.e. KQML messages). The technical staff can therefore perform the operations that the game engine will perform during a game session, namely creating modifying and removing instances or querying the state of the knowledge base. Moreover, the actual rules fired at each interaction can be monitored.

Figure 3 shows a snapshot of the graphical interface. The application is composed by five windows: (1) the “KB Box” window, above on the left, display the frame hierarchy and the instances created so far; (2) the “Control Box” window, displays detailed information on the selected element (i.e. either a frame, an instance, a message etc.); (3) the “Operation Packages Box” window, bottom on the left, stores the operations on instances already defined by not yet sent to the Evaluation Engine; (4) the “Retrievals” window, below middle, stores the queries to be submitted to the Evaluation Engine; and finally (5) the “Message Box” window, stores all the message sent to and received from the Evaluation Engine.

KBS actually interprets the KQML messages received by the Evaluation Engine and it maintains the consistency in the windows, in particular in the “KB Box” where the instances created, deleted and removed by either an user operation or the effect of a rule application are properly displayed. Yet we decided to maintain visible the mes-

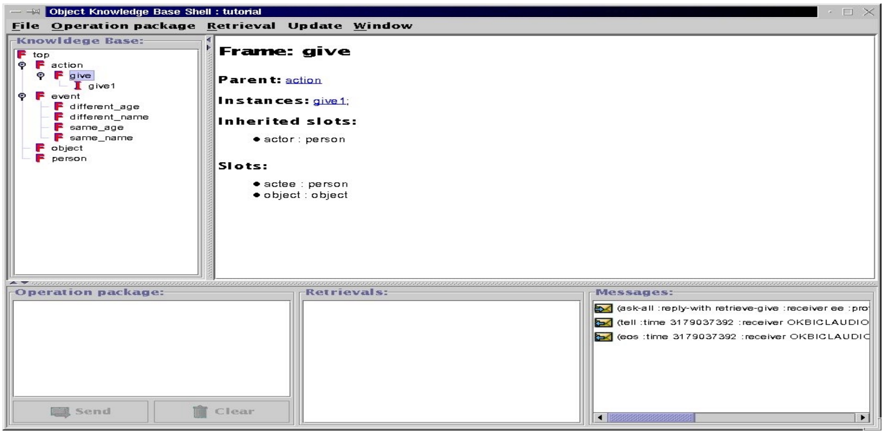


Fig. 3. The main window of the KBS application

sage exchanged to help the technical staff in better visualizing what is going on during a game session.

5 Conclusions

In this paper, we introduced a first attempt to build an authoring environment for the AI of (educational) computer games targeted to a technical editors staff. We think that in providing support of this kind of user testing is as much important as editing, in particular if the editorial works has to be made in parallel with the graphical design and with the programming, as it is usually the case.

This work is still in progress and it has been conduct in the context of a project funded by the European Community to develop a prototypical educational computer game, we would like to acknowledge the support of the other partner of the project for their suggestions and fruitful discussions.

References

1. Laird J. and van Lent M., Human-level AI's Killer Application: Interactive Computer Games, AAAI2000 Invited Talk, Austin TX, 2000.
2. Labrou Y., Semantics for an Agent Communication Language, PhD Thesis, University of Baltimore MA, 1997
3. Forgy RETE: a Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem", Artificial Intelligence, 19, 1982.
4. MacGregor. R. M., A Deductive Pattern Matcher, in Proceedings of the Seventh National Conference on Artificial Intelligence, (AAAI 88), pp. 403-408, 1988.

5. Yen, CLASP: Integrating Term Subsumption System and Production Systems" IEEE Transaction on Knowledge and Data Engineering, 3(1) 1991.
6. Newell A. ,Unified Theories of Cognition, Cambridge University Press. Cambridge MA, 1991.
7. Anderson J. R., Rules of the Mind, Lawrence Erlbaum Associated, Hillsdale NJ, 1993.

Generation and Implementation of Mixed-Reality, Narrative Performances Involving Robotic Actors

Nikitas M. Sgouros and Sophia Kousidou

Department of Informatics, University of Piraeus, Karaoli & Dimitriou 80, 18534 Piraeus, Greece
{sgouros, nova}@unipi.gr

Abstract. Recent advances in robotics and multimedia technologies have created new possibilities for staging narrative performances involving robotic actors. However, the implementation of these types of events is complicated by the lack of appropriate direction and execution environments that can deal with the complexity of these productions. This paper seeks to address this problem by describing CHOROS, a Java-based environment for authoring, direction and control of narrative performances. In particular, CHOROS allows the story author to annotate the performance script with stage directions. Furthermore, the environment offers to the performance director an augmented reality interface for planning the behavior of the actors. Finally, the system uses vision-based tracking methods and behavior-based control for adjusting the behavior of the robotic actors according to the director instructions during the performance.

1 Introduction

Recent advances in robotics and multimedia technologies have created new possibilities for integrating robotic actors in narrative performances. Such a development greatly enhances the means of expression available to creators of storytelling environments by allowing them to stage mixed reality events in which human and robotic actors along with various other multimedia objects strive to create immersive and enjoyable narrative experiences. Unfortunately, the development of these types of performances is hampered by the lack of appropriate directing and execution environments that can deal with the complexity of conceiving and the unpredictability of executing these projects. The majority of these productions uses a mixture of traditional multimedia authoring tools and robot programming environments that can deal only with isolated aspects of the development and execution process. In particular, traditional multimedia authoring and presentation tools can only describe the spatial and temporal relationships between the multimedia objects (i.e., video, audio, graphics) that comprise a multimedia application. These tools lack the ability to automatically track the behavior of human or robotic actors and associate it with the execution of various multimedia objects during the staging of a narrative event. On the other hand, current robot programming environments are not able to describe and execute

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 69–78, 2001.

behaviors that should be synchronized with the rendering of various multimedia objects (e.g., synthetic actors, speech, video or audio clips). In addition, the appearance and behavior of most current mobile robots is not expressive enough to support acting. Consequently, a new generation of development and execution environments for narrative performances featuring robotic actors needs to be developed so that:

- story authors can incorporate stage directions in the narrative text
- directors can plan off-line the behavior of all the actors in a performance
- actor behavior can be automatically tracked and guided during the staging of the performance according to the director plan
- the behavior of the robotic actors can be expressive enough in order to support acting.

This paper describes CHOROS, a development and execution environment for narrative performances featuring human and robotic actors that seeks to address these requirements. At its current stage of development the system provides assistance to the author and director of a narrative performance. In addition, it provides a run-time environment that monitors the implementation of the directing plan during the performance.

In particular, CHOROS allows the story author to incorporate stage directions in the performance script by annotating the spoken dialogue between the actors with prosodic information using the Java Speech Markup Language (JSML). In the case of the robotic actors this information is fed to a speech synthesizer that verbalizes appropriately the annotated text during the performance.

The environment offers to the story director an augmented reality interface for drawing the paths that will be followed by the performance actors. Planning is performed on a top-level view of the actual performance space as captured by an overhead video camera. We refer to this space as the *stage* for the event. The director can specify graphically the temporal constraints describing the use of synthetic, audio or video objects at specific regions in these paths by ordering these objects in *path-bound timelines*. These are timelines that are drawn in parallel with the actor paths on the stage view. We refer to the specified paths and timelines as the *director plan* for the narrative performance. The environment automatically analyzes the director plan to extract a set of both qualitative and numerical constraints on the spatial and temporal behavior of the actors during the execution of the application. The numerical constraints include the motion parameters for the robots (speed, acceleration etc.) that are theoretically necessary in order to follow their designated trajectories and synchronize their movement with the rest of the multimedia objects. The qualitative constraints include the groups of actor motions that should be executed concurrently during the performance based on their media constraints along with the spatial relations between them (e.g. parallel, converging, diverging etc.). CHOROS provides the director with a simulation environment in which s/he can visualize and monitor the execution of his/her plan in the event stage.

At run-time, all the original and extracted constraints are fed to an execution module that constantly tracks and adjusts the behavior of the robots in order to follow the director plan. Tracking is using frame-differencing operations on a grid-based decomposition of the stage view to detect the position of each actor in space and analyze the direction and speed of its motion. Adjustment seeks to deal with the unpredictability of controlling robot movement at run-time. This goal is achieved by mainly preserving the qualitative constraints between the behavior of the actors and the rest of the multimedia objects in the face of frequent deviations of the actors from their designated trajectories. These deviations are caused by either sensor or actuator errors on the robots or the approximate interpretation of the scenario by the human actors.

CHOROS can be used for the development of narrative or more general multimedia performances that involve robots such as robotic theatre productions or puppetry, interactive playspaces [1], dance productions, programming of tour-guiding or entertainment robots and creation of robotics-based special effects for the movies.

2 The Directing Process

Directing environments for storytelling environments involving robotic actors need to provide intuitive methods for describing the behavior of these actors. For this reason, the directing process in CHOROS uses a mixture of augmented reality and timeline-based techniques.

In particular, the system provides the director with an augmented reality environment in which the trajectory of each robot can be drawn on a top-level view of the stage. In this case, the path planned for each robot consists of a sequence of line segments. The director determines the starting and ending points for each segment by clicking on the desired points on the screen. These points can signify either a change in the direction of movement of an actor or the enactment of constraints associating the spatial location of the actor with the rendering state of various multimedia objects. In the second case, the director specifies the frame number of a video object or the time position of an audio clip that should be rendered whenever the robot reaches the particular point in its trajectory. In addition, the director can specify the piece of text that should be verbalized by the speech synthesizer whenever the robot reaches such a point. Since frequent sensor or actuator errors by the robots make it very difficult to achieve an exact synchronization of this sort at run-time, the director is able to specify a region in space centered on the specific point in which the particular constraint should be satisfied. Once such a region has been specified for both ends of a line segment, the system draws a path-bound timeline parallel to this segment. This timeline depicts the starting and ending positions along with a set of intermediate positions of the multimedia object that will be rendered while the actor follows the specific line segment. The existence of such a timeline provides to the director an effective way of visualizing the association between the behavior of the robots and the rendering of various multimedia objects. In addition, this timeline allows the developer to monitor effectively the synchronization between the actors and the multimedia objects at execution time.

For example, Figure 1 provides a snapshot of the stage view in CHOROS that contains two robotic actors. The stage floor in CHOROS is covered with a black material in order to facilitate real-time tracking of the behavior of the actors. The figure depicts the path planned for each actor and its associated timeline, which, in this case, refers to a video object for each actor. The rectangular areas in the path of each actor describe the association of a particular region in the trajectory of the actor with a specific rendering state of a multimedia object. In the figure, we refer to such an association as a media-driven constraint.

The resulting directing environment seeks to describe the major ways of associating actor behavior with the use of various multimedia objects in a performance. More specifically, such an association can be either *actor-driven* or *media-driven*. In the first case the behavior of the robots in the stage drives the use of the multimedia objects.

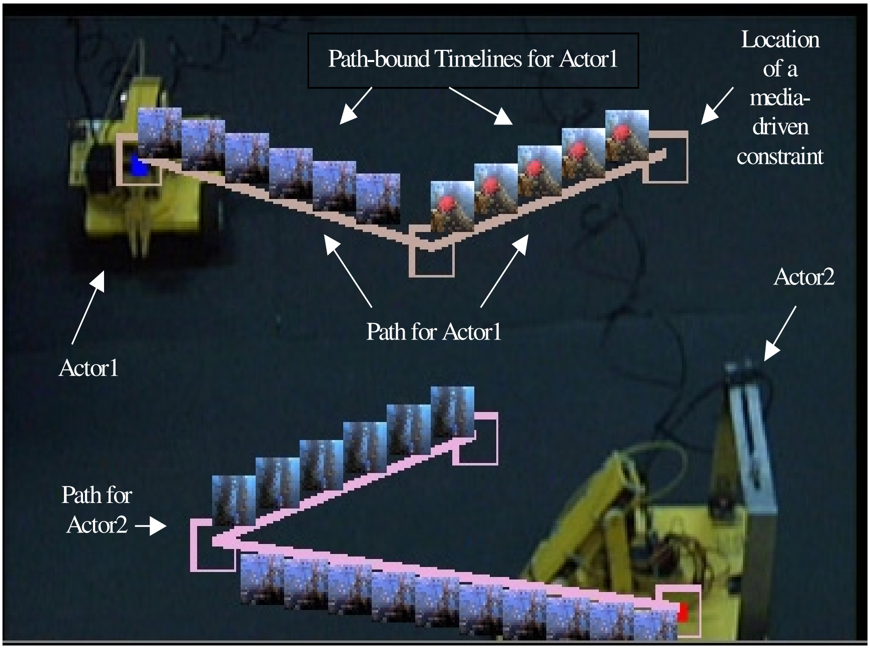


Fig. 1. The stage view in CHOROS.

For example, an actor in a robotic theater production can verbalize a particular text segment whenever it approaches certain areas in the stage. On the other hand, a media-driven association is one in which the execution of the multimedia objects dictates the behavior of the robotic actors. For example, the music in a dance performance usually forms the basis for organizing the movement of the robots participating in it.

2.1 Specification of Actor-Driven Associations

The directing process supports two types of actor-driven associations. The first one covers *location-specific* cases in which the execution of a set of multimedia objects starts or terminates when one or more robotic actors enter or exit from specific areas in the performance space. These areas can be either static, as in the case of the stage area in the robotic theater example, or mobile, such as the region surrounding another human or robotic actor during a performance. In the second case, speech, video or audio clips are triggered during the negotiation of the space that separates the actors. For example, a robot (e.g. R) might execute appropriate audio clips (e.g., sound a horn) that warn another actor (e.g. A) against approaching it. These clips are activated whenever A enters a region designated by the author that constantly surrounds R.

The director is able to describe location-specific interactions by indicating in the augmented reality environment specific areas in the performance space and associating the execution of multimedia objects with certain actors entering or exiting these areas. At the time of definition, the system checks whether these areas contain another actor. If so, the area becomes bound with the robot and follows it through space, otherwise the area is considered to be static.

The second type of association covers *behavior-specific* cases during which a particular sequence of commands executed by a robotic actor should trigger or terminate the execution of a set of multimedia objects. For example, when an artificial pet wags his tail a series of audio clips from real pets engaging in this behavior can be rendered in order to reinforce the illusion of a real pet. We are currently developing a graphical environment for associating sequences of motion commands of an actor to the enactment of constraints affecting the execution of multimedia objects. This environment will allow the director to associate reflex-like behaviors of the robots with appropriate multimedia objects.

2.2 Specification of Media-Driven Associations

The environment identifies automatically media-driven associations in the behavior of a particular actor (e.g. R) through the detection of *media-driven segments* in its path. A media-driven segment for R is a sequence of line segments that contain in its starting and final points a constraint on the use of the same multimedia object. For example if the author has specified that at point A in the path for actor R video object V should be in frame position F_1 and at point B in the same path V should be in frame position F_2 then the sequence of line segments in this path that start from A and end in B form a media-driven segment in the trajectory of R.

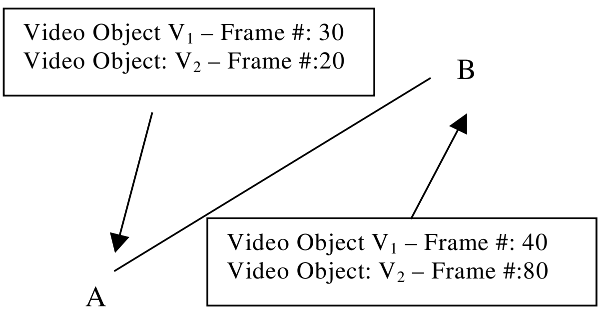


Fig. 2. Example of an inconsistent media-driven segment.

Once the system has detected a media-driven segment it checks whether it is *consistent*. A consistent media-driven segment for robot R is one for which there is a set of motion parameters (e.g. speed) for R that allows it to follow its designated trajectory in the segment and satisfy all the constraints at its starting and finishing points. For example, if we assume that video objects V_1 and V_2 should be rendered with the same frame rate then Figure 2 depicts an inconsistent media-driven segment because there is no speed for the actor that will allow it to traverse the segment and satisfy the constraints for both video objects V1 and V2. The system notifies the author of inconsistent line segments in order to take remedial action.

2.3 Qualitative Analysis of Media-Driven Segments

The analysis of media-driven segments seeks to extract a set of qualitative constraints that describe the spatial and temporal relations between the behavior of the actors during the performance. Analysis proceeds through the execution of the following sequence of steps:

1. Detection of concurrent points and segments in the actor paths.
2. Extraction of qualitative spatial constraints on concurrent segments.

Detection of concurrent segments. Two points in the paths for two actors (e.g. R_1 and R_2) are *concurrent* if they are extreme points (i.e., starting or final points) in their respective media-driven segments and they share at least one constraint.

For example, in Figure 3 the points A_1 and A_2 in the paths for actors R_1 and R_2 are concurrent since they share the same constraint for audio object C.

Two media-driven segments for two actors (e.g. R_1 and R_2) are *concurrent* if they have concurrent extreme points. For example, in Figure 3 the segments A_1B_1, A_2B_2 and A_3B_3 that correspond to actors R_1, R_2 and R_3, respectively, are concurrent because their extreme points A_1, A_2 and A_3 along with B_1, B_2 and B_3 are pair-wise concurrent.

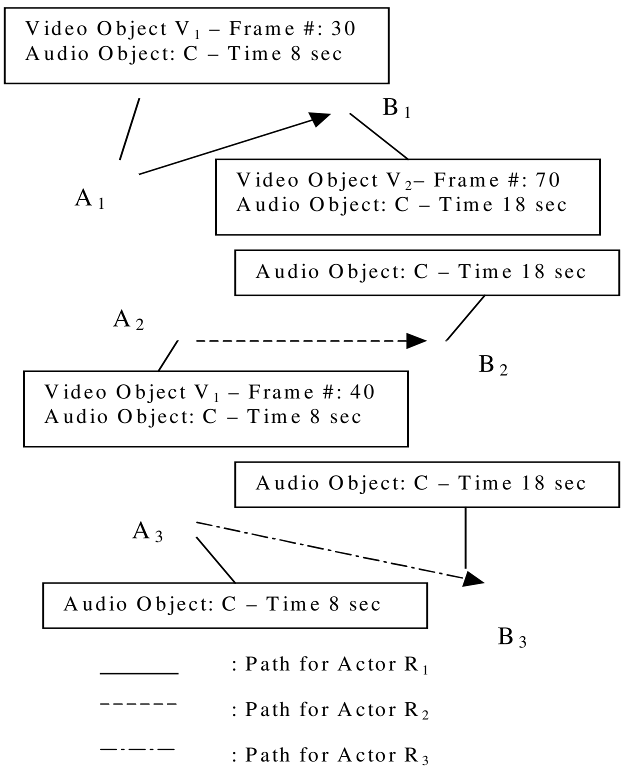


Fig. 3. Examples of concurrent points and segments in media-driven associations.

For each multimedia object the authoring environment lists the media-driven segments that are constrained by it. For each such object the set of media-driven segments that are pair-wise concurrent form a *concurrent set*. For example, in Figure 3 a concurrent set for audio object A is: { A_1B_1, A_2B_2, A_3B_3 }.

Extraction of qualitative spatial constraints. During media-driven interaction it is often the case that the lines and shapes each actor produces through space can be related to those of other actors through copying, complementing or contrasting relations. The establishment of these relations creates a momentary image which holds meaning for the audience. The purpose of this step is to extract a set of qualitative spatial constraints that describe the orchestration of the movement of all actors. Currently, the line segments in each concurrent set are classified as:

1. Parallel, if they have approximately the same slopes. Two parallel segments can be *opposite* if the actors involved move in opposite directions or *analogous* if they move in the same direction.
2. Converging, if the distance between their end points is less than a user-specified threshold while their starting points are further apart.
3. Diverging, if the distance between their starting points is less than a user-specified threshold while their end points are further apart.

3 The Execution Process

The execution process constantly tracks and adjusts the behavior of the robotic actors in order to follow the director plan. Tracking is using background separation and frame-differencing operations on a grid-based decomposition of the stage view to detect the position of each actor in space and determine the direction of its motion and its speed. Adjustment seeks to deal with the unpredictability of controlling robot motion at run-time. This goal is achieved by mainly preserving the qualitative constraints governing the behavior of the robots.

3.1 Actor Tracking & Guidance

The tracking process accepts as input the stage view and assumes that the background of this view will remain constant at run-time. This allows it to perform background separation at each frame and then compute the difference between successive frames. Each one of the resulting frames is then mapped to a grid of twelve cells. Each grid cell is assigned a number, which is equal to the number of interesting pixels in it, i.e., the pixels with values above a noise threshold. This threshold has been computed during a calibration procedure for the particular stage view. Furthermore, for each grid cell with a positive number, the method computes the center of gravity of its interesting pixels. In order to determine the location of each actor in the stage view the tracking process picks the grid cells with the highest numbers and uses a minimum distance classifier which assigns each center of gravity in these cells to a particular class that represents an individual robot. This particular tracking process has achieved a success rate of over 90% in CHOROS.

The results of the tracking process are fed to a guidance module that forces each robot to follow as closely as possible the current line segment in its path. In particular, this module constantly issues a set of motion commands to each robot that seek to minimize the distance of the current location of the actor from the final point of its current segment in the stage view. At each point in its trajectory these commands move the robot in a direction approximating the direction of the line connecting its current location in the stage view to the final point of the current line segment in the same view.

3.2 Behavior Adjustment

Behavior adjustment accepts as input the numerical and qualitative constraints describing the association between the behavior of the robots and the use of various multimedia objects. It seeks to preserve these associations by issuing appropriate motion commands to the robotic actors.

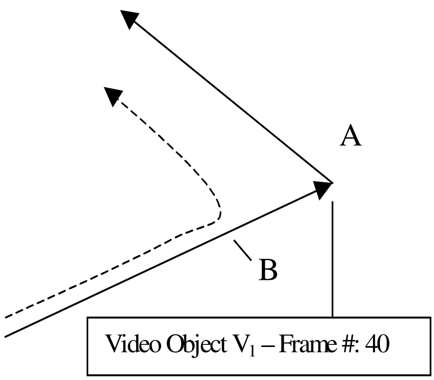


Fig. 4. Change the behavior of an actor as soon as the relevant constraint on a multimedia object is satisfied.

Because frequent sensor or actuator errors make it almost impossible to satisfy the numerical constraints on all segments of an actor path, the adjustment process seeks to preserve primarily the qualitative constraints associating the traversal of each segment with the rendering of various multimedia objects. To this end, the process applies the following rules:

Whenever a multimedia object reaches a rendering state that has been associated with a change in the behavior of an actor in a media-driven segment, then this change will take place irrespective of the location of the actor in the stage.

For example, Figure 4 depicts the desired and the actual trajectories that will be followed by an actor, assuming that video object V_1 reached frame 40 at point B and not at point A as it was specified during authoring. The desired trajectory is drawn with a continuous line, while the actual trajectory is drawn with a dotted one. In this case the actor will actually turn left at point B.

If an actor reaches a location in space before a multimedia object reaches a rendering state that has been associated with this location then the actor will remain in this location until the desired rendering state is reached.

In Figure 4, for example, if the actor reached point A before video object V_1 reached its 40^{th} frame then the actor will remain in A until frame 40 is rendered. It will then continue to follow its specified path.

Both rules ensure that concurrent media-driven segments will produce concurrent behaviors at run-time. Consequently, the temporal structure of the behavior of the actors that was prescribed during the authoring phase will be preserved. However, the application of the first rule will not preserve the exact spatial structure of the behavior of the actors. In order to preserve the qualitative constraints on the actor movements, the execution process applies the following rule:

If a group of robotic actors begins to follow a concurrent set of media-driven segments then the system will try to satisfy the qualitative spatial relations, if any, between the elements of this set. In particular, parallel segments must remain parallel, while convergence or divergence relations should be preserved between the elements of this set.

Consequently, if there are deviations between the actual starting position of a segment in the concurrent set and its prescribed starting position from the authoring phase, the system will determine a new final position for this segment. The computation of these new segments will take place in the stage view and it will try to satisfy the geometric relations between the elements of the set. The length of the new segments should not exceed the length of their respective segments from the authoring phase in order to ensure that the media constraints at their end points can be reached.

4 Implementation

CHOROS has been coded in Java using the Java Media Framework API for dealing with the multimedia objects and the Java Communications API for managing the robots. The robots used in CHOROS consist of a pair of low-cost mobile manipulator kits from Lynxmotion that communicate with a Pentium II PC using its two serial ports. The robots carry no sensors. The only sensor that is used by the system is an overhead video camera connected to the PC that provides a 320 x 240 stage view for the application.

Up to this date, CHOROS has been used in a series of trials implementing media-driven associations between the behavior of the robots and the rendering of speech, audio or video objects.

5 Conclusions, Related and Future Work

This paper describes a direction and execution environment for narrative performances featuring autonomous mobile robots. This work extends research on interactive playspaces [1] by allowing their integration with robotic actors and providing appropriate development environments for them. In addition, it seeks to support the creation of a new generation of authoring and execution systems for interactive narrative performances that coordinate the interaction between physical actors (i.e., robots or humans) based on higher level plot structures and/or audience reactions [2-3].

Future work in this area will focus on extending the means of expression of the director during the authoring phase. This will be achieved through the development of a rich vocabulary for composing movement and linking it with the rendering of various multimedia objects. Systems for analyzing and transcribing movement, such as the Laban or Benesh notations [4-5], can provide inspiration for implementing these types of extensions. Furthermore, future research will seek to implement expressive behaviors in the robotic actors that are suitable for acting.

In terms of content development, CHOROS is currently being used for the creation of a robotic theater production in conjunction with Yiannis Melanitis [6], an artist working on robotic performances. In this production, the system is used for planning and controlling the behavior of two robotic actors, a hexapod and a mobile manipuilator. In addition, CHOROS controls the movement of a pair of robotic cameras that move in the performance stage in order to capture the development of the event according to the director instructions and broadcast it on the Web.

Acknowledgements. This work was supported by the Greek Secretariat for Research and Technology under the program «Hygeiorobot» (GSRT-PENED 99).

References

1. Bobick, A. F. et al, The KidsRoom, *Communications of the ACM*, vol. 43, no. 3, March 2000.
2. Sgouros, N. M., Dynamic generation, management and resolution of interactive plots, *Artificial Intelligence*, 12(3), 29-62, Elsevier (1999).
3. Sgouros, N. M., Detection, Analysis and Rendering of Audience Reactions in Distributed Multimedia Performances,» *ACM Multimedia 2000*, Los Angeles, CA, November 2000.
4. Laban, R., *Principles of Dance and Movement Notation*, New York: Dance Horizons London: MacDonald & Evans, 1996.
5. Blom, L. A., Chaplin, L. T., *The Intimate Act of Choreography*, Dance Books Ltd., London UK, 1989.
6. http://digitalart.asfa.gr/students/melanitis

Film and the Development of Interactive Narrative

Andy Clarke and Grethe Mitchell

Kinonet, 78 West Kensington Court, Edith Villas, London W14 9AB
andy@kinonet.com
Dept. of Innovation Studies, University of East London,
Docklands Campus, 4-6 University Way, London E16 2RD
grethe@kinonet.com, grethe@uel.ac.uk

Abstract. This paper explores narration in film and in videogames/virtual environments/interactive narratives. Particular attention is given to their use of the continuity of time, space and action and this is used as a means of classifying different types of work. The authors argue that the the creators of these videogames etc. need to have more authorial presence and that this can only be done through abandonning their traditional reliance on the continuity of time, space and action.

1 Introduction

There is a clear crossover between film and interactive computer-based entertainment. This crossover occurs in a variety of forms and at a number of levels, and is appropriate given film's central role as a storytelling medium and the clear formal similarities that exist between films and computer-based media - the main ones being that they are both screen-based, they are both time-based and they both convey most of their information visually.

If we take videogames as an example, we can see that they borrow the established conventions and iconography of film. The guns in Unreal Tournament look like those in Aliens because doing so provides the designers of the game with a shorthand to describing the characteristics of each weapon. Similarly, the lighting, camera angles, and music of Resident Evil are like those in George A. Romero's Living Dead series of movies because drawing upon the conventions of the horror movie genre (and this subgenre) provides the game with a shortcut to creating a sinister atmosphere.

But what is more significant than this is the way in which a game such as Soul Blade uses its "virtual camera" - tracking and zooming to follow the fighters, and slowly tracking back and up to give the impression of the life draining from your character's body when you die - or the way that Resident Evil uses editing. In both of these examples, the videogame is using the language of film, but in a way that is subtly different to how it is used in film.

In other papers, we have looked in detail at various formal aspects of the videogame, with a particular interest in the relationship between viewpoint and identification, and between immersion and narrative in videogames[1, 2]. Although there is some benefit in choosing another formal aspect of film and virtual

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 81–89, 2001.

environments to explore here or to examine in fine detail the way in which a single one of these environments presents its narrative, both of these tasks are probably best left to a book-length examination of the subject. In stead, we intend to pursue another option in this paper, to draw these - and other - threads together to explore a single theme: narration.

Narration is something very distinct from narrative. The narrative is the story: “what happened”. The narration, on the other hand, is the storytelling. As we will see in the next section, there are elements of film narration that are common to all films, no matter what they are about (though each film may use these narrative techniques in a different way, possibly using some not at all). Are there a similar set of narrative techniques in interactive narratives, and if so, what are they? Are they complete, unified and coherent, and if not, where are the lacunae and what do these absences and omissions tell us about the nature of the medium and/or the stories that people choose to tell in it? Are there better ways to present interactive narratives or to present narratives in interactive environments?

It is worth pointing out that for reasons we explain later in the paper, we make little or no distinction between the various types of interactive narratives, videogames, arcade games, immersive environments, and so on. As a result, we will use these terms somewhat interchangeably, and our examples will be drawn mainly from the world of videogames as these are the most widely available, should the reader wish to view the work for themselves.

2 Narration in Film

Film has, over the years developed a wide range of formal techniques (formal in the sense of dealing with film form, rather than in the sense of being a precise, coherent, complete or logical system). These conventions are so familiar to us that they become “invisible” - they are helped in this respect by the fact that film normally engages us so strongly at an emotional level, as a narrative, that it is difficult to identify at the same time the techniques by which it is presenting this narrative or having this emotional effect.

It is useful therefore to list the most important of these formal components (particularly for the benefit of those who have not studied film theory).[1] They can be summarised as follows: the *mise en scene* (the choice of actors, costumes, props and setting; the blocking of the scene; the use of special effects; etc.); the shot (the choice of camera position, focal length, film stock, aspect ratio and framing; the use of a moving camera or zoom lens; the choice of lighting style, setup and level; etc.); the juxtaposition of shots (the speed and style of editing; the use - or non-use - of continuity editing; the use of the long take and/or a moving camera as an alternative to editing; the use of optical effects such as dissolves between shots; etc.); sound (the use of sound, sound effects, music and silence; etc.).

This list is not meant to be complete, definitive or exhaustive, but is intended merely to give some indication of the type of elements involved and highlight some of the more important. There is clearly a great deal of overlap between categories and ongoing discussion about how to divide even the partial list of properties above between them - one could, for instance, easily class lighting as part of the *mise en scene*, rather than as part of the shot as the lighting is often motivated by the choice of

[1] Good introductory texts are Film Art: An Introduction [3] and The Cinema as Art [4].

location and the lights which are present in the shot. The use of a moving camera or zoom lens is also interesting in that it is normally the use of this within a long take that replaces editing, rather than the long take per se.

One can likewise easily add to this list of categories or create subtle distinctions within them. One example of this is within sound - between sound that comes from things which are (or could be) present in the scene (diegetic sound) and sound that is not motivated in this way (non-diegetic). Even this can be further subdivided between diegetic sound from onscreen people or things, and that from offscreen ones.[2]

But these formal elements of film cannot work without the filmmaker and their audience having an agreed understanding of what they "mean". There is an infinite number of ways that the filmmaker can choose to take a shot (different camera positions, focal length, etc.) but these choices are meaningless unless there are conventions that give meaning to the various possibilities - i.e. that says that a wide shot from high up looking down on a figure means one thing, and a low tight shot looking up at the same figure means something else.

In fact, it is not as clear cut as this. Each of these shots can have a whole range of different meanings according to its context (both on the basis of its position within the narrative and depending on what shots precede and follow it). Indeed, the filmmaker can create new meanings through the context of the shot and these can be incorporated into the evolving body of film grammar.

What all of the above should indicate is that when we use terms such as "film language", "the conventions of film" and "film grammar", we are not talking about a single, fixed, consistent, coherent set of rules that can be written down - indeed, attempts such as Metz's Grande Syntagmatique[3] have failed to define an internally consistent set of rules of film grammar even within a single film.

It should therefore be obvious that when we talk about applying the lessons learnt from film theory and practice to the creation of virtual worlds and interactive narratives, we are not talking about distilling the conventions of film into a neat set of axioms and formulae which then can be programmed into the computer - neither films nor virtual environments represent bodies of work which are consistent and coherent enough in terms of form, content, aims, resources or techniques to allow this.

But in spite of this, we still regard the theories and techniques of film as being potentially far more useful than those of other media. There are a number of reasons for this, but the key one is that they can often be applied both at a "macro" level and at a "micro" level. To clarify what we mean by this, compare, for example, Propp's theories of character and plot structure [6] to those of character, genre and *mise en scene* in the Western. Propp's theories describe the role that the villain has in the narrative, but offer very little assistance in how to implement them in an interactive narrative. There is no advice as to how this villain should look, how he behaves (beyond the broadest description), and so on.

Film theory offers far more concrete guidance in this respect. The conventions of genre sketch out the role that the villain has in the narrative (as in Propp), but there is a whole host of other advice that it (and other film theories) can give us to "flesh out" this character. They will tell us that the villain will wear a black hat and will be badly shaven. In addition, it will provide a set of generic locations for the story to take place

---

[2] This is further complicated by the fact that the same sound may be diegetic in one shot and non-diegetic in the next.

[3] See Stam, Burgoyne and Flitterman-Lewis [5] for an overview of Metz's theories and their weaknesses.

around (jail, saloon, main street, etc.) and generic set-piece events for the characters to engage in (crooked card game, quick-draw shoot-out, and so on).

But it doesn't stop here. The theories of film also provide some guidance as to how to best convey the information above to the player. The conventions of the Western are that the action is often shown in a wide shot, so this can be our default way to render a scene, but the fact that the villain is badly shaven is best conveyed by showing him in a close-up. Likewise, other objects - such as the sheriff's star - are known to be significant for this genre and will likewise be worthy of close-ups.

This has hopefully given some brief indication of how a (silent) interactive Western system could be assembled from what we know about the Western from film theory.[4] Such a system is far more concrete that the idealised storytelling engines in Aaseth [7] and Laurel [8].

3 Narrative and Narration in Interactive Entertainment

Many writers including Laurel [9], Murray [10], Aarseth [11], Poole [12], and others have written about interactive narrative, and although different writers may use different terms, most come to a common conclusion - that the more freedom the player/user has to intervene in the narrative or choose their own path through the narrative, the weaker the voice of the author becomes.

We would argue that the problem is not the player's intervention in the narrative, but rather their intervention in the narration. The clear and undeniable problems that current interactive narratives of all types have - a lack of empathy with the characters, a lack of engagement with the events of the story - has more to do with the narration of these works than with their narratives: the most exciting of stories can be made dull when presented in an uninspiring way, while the plainest of events can be made interesting and exciting through its presentation.

Nowhere is this more true than in videogames. At a surface level, these are the most seductive of stories to immerse oneself in - they tell action movie stories with glossy computer graphics and put you in control of the hero of the story - but as you play them, there is a sense of disonnectedness: even though you may have the sense of "being there" and jump (in the real world) when you are ambushed (in the game), there is little or no sense of engagement with the narrative or with the characters within it.

Film narration, as outlined above, is the interplay of a set of conventions regarding the presentation of people, objects and events onscreen. We must therefore ask whether a similar, parallel or equivalent set of conventions exists - or is developing - within interactive narratives, and whether the use of these techniques can solve the problems mentioned above and increase the emotional impact of these works.

As we have indicated before, there are clear formal similarities between film and computer-based forms of entertainment that have allowed many aspects of film to already be adopted by these media. Most of these similarities are so obvious that they are barely noticed and only a brief run-though of them is needed: they are both screen-based media; they are both time-based media; they both use technology in their production and presentation; they both tell their stories predominantly though image, rather than through dialogue.

---

[4] Creating characters that speak and respond to speech is a separate AI issue, though again this task may also be simplified by the conventions of genre.

There are, however, also differences. The most fundamental of these - if one leaves aside, for the moment, issues such as the scale of the cinema screen or the fact that films are projected - is the nature of view and viewpoint in the two media. A film, on the whole, uses a variety of shots to tell its story and while some of these shots may coincide with the viewpoint of one of the characters, most of the time they don't. The filmmaker decides what object, person or event to show on screen - and how to show it - and the viewer of the film can only choose where to focus their attention within this limited view given to them (or to look away).

In interactive environments, however, the player tends to have a continuous view of the action, typically through the "eyes" of their character or from behind them. Taken individually, neither of these points - having a continuous view or seeing through the "eyes" of the character - is a problem. The problem comes when they are used exclusively and thus impose a continuity of time, space and action on the videogame: the designer of the world is left with little opportunity to control what the user sees or how they see it.

Under such conditions, the designer of a virtual environment cannot prioritise an object through showing it in close-up as a filmmaker would do - they must do it through the design of that object, its placement in the world, the way that it is lit, and so on. They are likewise forced to create mood through the design of the world, the lighting, the props and decor, etc. Essentially, they are reduced to one narrative technique: that which we refer to in film as *mise en scene*.[5]

But there is a limit to how much the designer can prioritise an object solely through its design, placement and lighting. Virtual worlds therefore tend to have every non-essential item removed so that the few objects that you can interact with stand out enough to be noticed. Significant objects are also often constructed in a deliberately non-realistic fashion so that they stand out from their environment - keys in Quake, for instance, are oversized, floating in mid-air and rotating so that they cannot be missed. In effect, these videogames use the symbol of a key, or something that stands for the function of a key, rather than a key *per se*.

Virtual environments are sparsely decorated and furnished, present few objects to interact with, present them in a deliberately unreal way; they are also - for different reasons - sparsely populated (usually with characters that you can't talk to). The worlds themselves often have a maze-like structure, with paths blocked off until certain tasks have been done, as this is the easiest and safest way to guide the user.

The design of the world is, to a very real extent, the design of the narrative, and there are limits, therefore, to how subtle and sophisticated this narrative can be. As we have mentioned elsewhere [13], these are not problems that can be solved solely through achieving photorealistic quality in the rendering or greater realism in the animation.

4 The Continuity of Time, Space, and Action

In the previous section, we said that the continuity of time, space and action presented a problem in interactive narratives, (limiting the narration to a sparse *mise en scene* and the problematic design and placement of significant objects), and it is now necessary to explore this issue, its origins, and its implications in greater detail.

---

[5] Genre is another element, though through providing a ready-made set of characters, objects, locations, etc., it contributes greatly to the *mise en scene*.

By "continuity of time, space and action", we mean that the user experiences the environment as a space that exists as a whole independent of their presence or actions, and that the user's actions in this world are both presented and experienced as a single continuous event.

This is a complicated explanation for what is a very simple concept - that the world of the game is a world that we experience like the real world. At first, it may be difficult to imagine any alternative to this, but consider the following examples. In a film, for instance, scenes are made up of shots - taken from a variety of angles - edited together. This allows the film to jump backwards and forwards in time, or from one location to another. A play will likewise include jumps in time and between places not only between scenes, but also within them - indeed such jumps can take place mid-sentence. In addition, one can see how cubist paintings present multiple viewpoints in a single image, or different moments in time, or both.

What this should indicate is that there is not one form of presenting time and space that is "natural" and others that are "conventions" - they are all conventions. The fact that virtual environments tend to obey the continuity of time, space and action is not "natural" - it is merely a convention (and one that is particularly strange given the discontinuity of the player's experience: the fact that they will die within the game and be "reborn", that they will pause and resume the game, and so on).[6]

5 "Showing" and "Telling"

Interesting patterns emerge if one groups together works - both interactive and non-interactive - that use the continuity of time, space and action and ones that don't. There should be no surprise that this division (between works that use/obey the continuity of time, space and action and those that don't) means that most immersive virtual environments - whether they were produced as arcade games, research projects, simulations, interactive art, etc. - can be grouped together as they all obey the convention of continuity of time, space, action.

Also included this group are videogames such as Tomb Raider or Quake. We ignore the so-called "cut scenes" that may appear between levels in games such as these or their very occasional cut-away shots (such as to show what effect pulling a lever has had). The reason for this is that these are used so rarely that they do not interrupt the overall continuity of the player's experience to any significant degree: a player may spend an hour completing a level and only get a cut scene at the start and end of it; they will likewise only get a handful of cutaway shots.

A videogame such as Myst, however, doesn't fall into this grouping. This is because when you click to "walk" ahead in Myst, you don't see your viewpoint change as you take every step through the environment - you cut to what you would see from this new position. This contrasts with the experience of playing Quake, for example, where you viewpoint changes fluidly as you walk forward, or Tomb Raider, where the camera is constantly and continuously on your character.[7]

Also outside the grouping are text-based MUDs and MOOs. You might think that MUDs and MOOs would be grouped with the other virtual environments above, but

[6] One cause of this is the way in which virtual reality was "colonised" by architects at a very early stage - see Benedikt [13] for a prime example of this.

[7] Myst is a prime example - and in many ways the apotheosis - of the sparseness typical of virtual environments.

they are not. This is not because they allow actions such as "teleporting" or "emoting" (although both of these are interesting in that they are like cutting out your movement from one place to another - as you would do in a film - or stepping out of the narrative to deliver an aside - like an actor on the stage). In stead, it is because one can imagine that in a MUD, the world could be built so that when you go up a flight of stairs, you type "up" once to go halfway up and once more to reach the top, but when you go down, you can do so in a single "down". Here we are achieving the same sort of fluidity of time and space - and ability to compress and expand time and distances for narrative or emotional effect - that we have in film.

By dividing works of all types into those that obey the continuity of time, space and action and those that don't, one ends up with the following broad categorisations. In the category of those that obey continuity of time, space and action, we would place most modern videogames (including Quake, Tomb Raider, etc.), immersive virtual environments, multi-user worlds (such as Alphaworld), etc. The list of those that do not would include most old videogames, MUDs, MOOs, films, novels, plays, etc.

The pattern that emerges as a result of this distinction is easy to see, and echoes the terms used by other writers on (non-interactive) narrative: i.e. the distinction between "showing" and "telling", "mimesis" and "diegesis", "imitation" and "presentation. Essentially, what we are talking about here (with regard to interactive narratives) is the ability to structure the way that the events within the narrative are viewed or experienced by the user, rather than seeing the limit of one's role as "author" of the world as simply to decide what happens within it.

6 The "Birth" of Immersion

An activity such as the one outlined above allows us, on one hand, to identify similarities between various works that might, at first glance, seem to be very different and, on the other, to perceive differences between those which might initially appear to be similar.

It has other benefits too. If one examines the history of videogames in the light of the distinction between works that obey the continuity of time, space and action and those that do not, other interesting points emerge. One comes to realise that it is only relatively recently that the videogames obeying this continuity of time, space and action have come to dominate as they do now - before this the majority did not obey this convention.

If one looks at 1980, for example, one can see that virtually all of the top arcade games - Defender, Missile Command, PacMan, Space Invaders, Asteroids - consisted of separate, distinct levels of stylised play within a game area that is typically no larger than the screen. Only Battlezone stands out as an exception to this rule: here the action - tank combat - takes place in a "real" space, rendered with realistic 3D perspective (albeit with only vector graphics).[8]

If one looks at the current state of computer, console and arcade games, however, one sees the situation reversed: they are all dominated by games which *obey* the continuity of time, space and action. The crossover point is difficult to place precisely, but 1992/93 is a critical moment, with the release of first-person "shoot-em-ups" such

---

[8] PacMan etc. do not class as obeying the continuity of time, space and action as their levels are short - we are talking about the *sustained* continuity of time, space and action.

as Doom. Before Doom, immersion was not essential; after it, immersion has become - to a very great extent - the *sine qua non* of a videogame.

7 Conclusions

From this brief examination of narration in film and interactive narratives, it is possible to draw some conclusions about the way that interactive narratives can develop, and to do this, we will return to the distinction that we made between works that obey the continuity of time, space and action and those that do not. It is important to remember, however, that we are not making predictions about the narratives of these works - we are engaged in predicting developments in the *narration* of these works.

As we have pointed out earlier in this paper (and elsewhere), the continuity of time, space and action offers a quick and easy sense of immersion, but places severe limitations on the type of interaction that can take place and the level of narration that the creator of the game can impose on the events within the world and on the player's actions.

During the early nineties, at around the time that Doom was released, there was a fundamental shift in the design of videogames - from the games before this, which tended to not to obey the continuity of time, space and action, to those after it, which did. We can therefore speculate as to what videogames may have been like if they had continued along their original path and had taken advantage of the phenomenal increases in processor speed and computer graphics that have taken place since 1993.

Games were, at that point, already establishing conventions of their own for bridging space and time. A game such as Elite, for example, allowed the player to skip forward in time and cut out the boring bits of space travel (thereby progressing rapidly from one dogfight to another or one planet to another). Streetfighter, likewise, was establishing conventions of how to use montage - in a very comic book-like fashion - to convey the power of the fighter's blows, etc.

We believe that, over time, these games would have borrowed more extensively from the narrative techniques of film as these offer a ready-made and highly familiar set of conventions that they would have been able to exploit fully through their increasing ability to perform real-time rendering.

We can imagine, therefore, that these games - and the players of them - would have been comfortable with a whole range of filmic techniques that break up the continuity of time, space and action for narrative ends. These would include: changing viewpoint in mid-action; bridging time or space with a cut; cross-cutting between separate locations, characters or threads of the narrative; incorporating flashbacks and flashforwards; and so on.

We therefore believe that the designers of videogames and interactive narratives should abandon their obsession with immersion and return to exploring the freedom that breaking the continuity of time, space and action gives them as narrators.

It is important to realise that what we are arguing for is a clear and distinct break with the dominant form of videogames and interactive environments (i.e. those that obey the continuity of time, space and action). It is not sufficient to keep producing this type of game, merely with a little film language added on.

Games such as Quake or Unreal Tournament do what they do extremely well - they offer a very strong sense of immersion and are very exciting to play. Film language has only a limited role to play in videogames such as these, and adding it inappropriately would detract from their strengths. But there is, however, a separate but equally valid form of videogame for which the narrative techniques used in film are ideal: those which aim for a sense of engagement in the narrative and/or the characters but which are currently unable to produce this because they obey the continuity of time, space and action.

We therefore believe that there needs to be a rediscovery of the “path not taken” in the design of videogames, virtual environments and interactive narratives - a return to the freedom that breaking the continuity of time, space and action provides. Essentially this provides the opportunity to *author* a narrative, rather than just design a world - the opportunity to choose how the events that happen in the world will be experienced, rather than just deciding what events to show.

We are talking here about reclaiming the virtual world as a storytelling medium, rescuing it from architects and those who would make it just a perfect, yet pale, imitation of the real world. We can currently only speculate on what exciting form these works will take.

References

1. Clarke, A., Mitchell, M.: Playing with Film Language. In: Coulter-Smith (ed.): The Visual-Narrative Matrix. Southampton Institute (2000)
2. Clarke, A., Mitchell, G.: Screen Play: Film and the Future of Interactive Entertainment. In: Earnshaw, R. A. (ed.): Digital Content Creation. Springer-Verlag (2001)
3. Bordwell, D., Thompson, K.: Film Art: An Introduction. Alfred A. Knopf (1986)
4. Stephenson, R., Phelps, G.: The Cinema as Art. Penguin (1989)
5. Stam, R., Burgoyne, R., Flitterman-Lewis, S.: New Vocabularies in Film Semiotics. Routledge (1998) 33-68
6. Propp, V.: Morphology of the Folktale. University of Texas Press (1968)
7. Aarseth, E.: Cybertext. John Hopkins University Press (1997) 104
8. Laurel, B.: Computers as Theatre. Addison Wesley (1993) 138
9. Laurel, B.: Computers as Theatre. Addison Wesley (1993)
10. Murray, J.H.: Hamlet on the Holdeck. M.I.T. Press (1997)
11. Aarseth, E.: Cybertext. John Hopkins University Press (1997)
12. Poole, S.: Trigger Happy. Fourth Estate (2000)
13. Clarke, A., Mitchell, G.: Screen Play: Film and the Future of Interactive Entertainment. In: Earnshaw, R. A. (ed.): Digital Content Creation. Springer-Verlag (2001)
14. Murray, J.H.: Hamlet on the Holdeck. M.I.T. Press (1997) 48-49,120-121
15. Benedikt, M. (ed.): Cyberspace: First Steps. M.I.T. Press (1994)

Virtual Storytelling as Narrative Potential: Towards an Ecology of Narrative

Clive Fencott

School of Computing and Mathematics, University of Teesside, Borough Road, Middlesbrough TS1 3BA, England
p.c.fencott@tees.ac.uk

Abstract. In this paper virtual storytelling is considered as narrative potential – the integration of agency and narrative. To facilitate this, an aesthetics of VEs is introduced as the context for the analysis of a popular computer role playing game. The game is analysed in terms of Perceptual Opportunities – a content model for virtual environments. From this analysis some inferences are drawn concerning the way in which agency and narrative may be successfully integrated to facilitate virtual storytelling.

1 Introduction

Storytelling is an ancient and venerable art which humans have subsumed to a variety of media for a variety of purposes. Equally ancient and as important is the enactment of ritual where attendance and participation are of primary importance. The huge and continuing marked for the novel and attendances at football matches are testaments to the power which both hold to this day. Two important technological manifestations of these would seem to be the feature film and the computer game. In the former we have the dominance of narrative over participation, while in the latter we seem to have the dominance of participation over narrative. In the Star Wars films we follow the story of Luke Skywalker - the Jedi Knight - but in the game Jedi Knight we become a Jedi Knight with our own, less grandiose stories to tell. The difference is important for virtual storytelling because we are essentially offering the user the potential to find and tell his or her own story - not necessarily the author's.

Various authors have stated the intuition that computer games and storytelling are mutually exclusive - essentially expressing the view that agency and narrative are irreconcilable, for example [1]. Despite these apparent difficulties, various forms of virtual storytelling have been proposed from a range of perspective, e.g. a more literary approach [2] and in terms of such concepts as interactive cinema [3]. This paper considers virtual storytelling from the point of view of a playable character in a Virtual Environment (VE) and views virtual storytelling as a form of narrative potential seen as a balancing of both narrative and participation and that therefore the two can co-exist in the right circumstances.

In order to better understand narrative potential it is first discussed in terms of an aesthetics applicable to Virtual Environments (VEs) in general. To support this and to provide a practical context Perceptual Opportunities (POs) - a general content model

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 90–99, 2001.

for VEs - are introduced and their applications to virtual storytelling in particular discussed. POs can be organised into perceptual maps that offer an alternative to storyboarding and are an excellent technique for virtual storytelling design and analysis. The application of POs to gain insight into the nature of virtual storytelling will be achieved through the analysis of the computer game Shenmue[1]. From this analysis some generalisations will be made.

First, however, some brief thoughts on the nature of narrative and why 'new media' and VEs in particular are different.

2 The Nature of Narrative

As Barthes [4] points our, narrative, in a diverse range of forms and to suit a diverse range of purposes, pervades human culture across the ages. Narrative also seems to be able to adapt to and make its home in almost any communications medium. Yet characterising narrative is no simple task and yet we need to do so if we are to consider what we might mean by virtual storytelling. At one level we could characterise narrative in terms of genre, plot, characterisation and connotation. If a VE possessed some or all of these it might well be considered an example of virtual storytelling. In a more structural approach, Roland Barthes [4] identifies the following as defining characteristics of narrative:

- Levels of meaning, i.e. basic units, the level of actions, the level of discourse, as well as the linear development of the narrative structure.
- A confusion of consequence - what is caused by what - and consecution – what simple follows what.
- If the narrative arrives at a major turning point it will always seek to choose the option which will prolong its life.
- Time is relative to the narrative logic – real time does not exist in narrative.

Of course this is not the whole story (pun intended) but it will allow us to address a particular question. How does virtual storytelling differ from narrative in general? If the answer to this is that it does not then there is nothing to investigate here. This paper assumes the answer is that it does and sets out to identify some of the differences.

3 Work and Meaning

When we read a book we usually don't consider the work we are doing to facilitate meaning. We might be aware of the fact that we are transforming our perceptions of abstract symbols into words and phrases - lexia in Barthes's terminology [5] - and then into meanings. But are we very often aware of the work of holding the book appropriately, of turning the pages, of aligning our head appropriately, of moving our eyes and focusing our eyes? The work of reading is so closely allied with the construction of meaning from written texts that we usually don't notice we are

[1] Shemnue and its characters are trademarks of Saga of America Inc.

performing it. When using VEs we will always have to exercise conscious work in order to find meaning in them. Moreover, this conscious work is at the heart of the pleasure of VEs and replaces meaning almost completely in some cases. Tetris is a good example where the actual meaning of the game objects - essentially simplistic jigsaw puzzle pieces - is far less important than the work of reconfiguring them into winning combinations. The latter is the driving pleasure of Tetris.

Aarseth, for instance, has proposed the idea of textual machines to capture the relationship between users and interactive digital media applications such as VEs [6]. Experimental results are beginning to show the degree to which perceived effort overrides the potential benefit of finding meaning [7]. Thus, one of the principle objectives of VE design is to implicitly suggest to people how they might organise the work of configuration in order to construct appropriate meaning. For virtual storytelling this should mean that such work is always subservient to the construction of meaning. In other words, the work of constructing configurations is never more interesting that the meaning of these configurations. But the work of meaning, participation, is at the heart of virtual storytelling just as much as it is of VEs in general and is one of the great pleasures that make the medium so engrossing.

4 The Aesthetics of VEs

Aesthetics gives us insights into the particular pleasures communications media offer and thus help us focus on designing content best suited to a particular medium. In general it will be the case that some of the aesthetic pleasures of a medium will be common to others but there will also be some that are particular to it. VEs are no different. However, because it is a relatively new medium the aesthetics of VEs are not well understood or documented.

However, the following is what might be called the Church-Murray aesthetics of VEs [8] because it is primarily a combination of Janet Murray's aesthetics of interactive digital media [2] and Doug Church's 'Formal Abstract Design Tools' for computer games [9]. It also draws on other work on presence, co-presence and transformation. The Church-Murray aesthetics, which has been found useful in both the teaching and design of VEs, consists of:

- Agency - the sense of feeling, at least to some extent, in control is composed of:
  - Intention - the formulation of goals and plans of action
  - Perceivable consequence - seeing the VE change as a result of intentions put into practice
- Narrative potential - the accumulation of meaningful experience as a result of agency - allows users to construct their own appropriate narratives. Narrative potential thus arises from agency but is not determined by it. NB. Some commentators place story or narrative here but for reasons stated in the introduction that would seem inappropriate when viewed in juxtaposition with agency. See below for further discussion.
- Transformation - refers to the ability of VEs to, temporarily, offer users new skills and powers or even to allow them to become different people or different species entirely.

- Presence and co-presence - which refer to users' senses of not only being in a mediated environment but also being present there with others.

This characterisation of the aesthetics of VEs is not definitive and could also include the pleasure of learning how to succeed in a VE. If we consider agency in relation to various applications domains then we see that the form of agency built into a virtual shopping mall must be quite different to that built into a computer game *shoot-em-up*. In the former, shoppers must feel in complete control if they are to spend money by giving up personal credit card details. In the latter, control must be partial or the game will not be of interest for very long.

In terms of narrative potential we can comment that all human situations have narrative potential in the sense that a good storyteller could find something in a trip to the corner shop to buy a bottle of milk the basis for a good narrative. The narrative potential in shoot-em-ups is typically of first I did this, then I did that, type. The sort of narrative potential we are looking for would be far richer and far more in accordance with the notions of genre, plot and layers of meaning, for instance, introduced above. It would be the sort of narrative potential that would lead users to realise the types of rich literary experiences that VE authors intended for them rather than a mere recounting of events.

Having identified an aesthetics of VEs, in particular agency and narrative potential, we now proceed to find the mechanisms that enable them.

5 Perceptual Modeling

The Perceptual Opportunities (PO) model of the content of VEs consists of a set of syntactic categories, which can be seen as attributes of any object that might conceivably be placed in a VE [10]. These attributes specify the way in which the object is intended to function in terms of communication. The syntactic categories into which POs can be characterised identify their role in achieving purpose and it is their planned interaction that gives us the overall structure we are looking for. We might thus see POs as a possible characterisation of the *lexia*, the base units, of virtual content rather than its scene graph representation. Figure 1 (below) shows how the range of POs can be broken down into three principle forms that are briefly discussed below.

At the first level we bread POs down to sureties, which deliver basic belief in a VE, surprises, which deliver the conscious purpose of a VE, and shocks, which are perceptual bugs that tend to emphasise the mediated nature of the VE. For a fuller discussion of POs and wider VE design issues see [8 & 10]. In this paper we are going to concentrate on surprises and the way in which they relate to agency and narrative potential.

Surprises come in three basic types:

- Attractors, which, as their name suggests, are designed to attract people's attention to possibilities for agency and should stimulate goal formation.
- Connectors are concerned with planning to achieve goals and supporting their attainment.

Fig. 1. A Characterisation of Perceptual Opportunities

- Rewards again, as their name suggests, should reward people for the exercise of agency.

Attractors, connectors and rewards can be grouped in triples and each triple will characterise a basic unit of agency. Such units seeking to identify what might stimulate the formulation of a goal, what work an planning is required to achieve that goal and what rewards are on offer for all this effort. A perceptual map is a loosely grammatical structuring of POs that seeks to ensure that users construct an appropriate temporal ordering over their attentions and activities within the VE. The simplest way of representing a perceptual map is by means of a table in the following manner:

Table 1. A partial perceptual map for a typical *shoot-em-up*

Attractors	Connectors	Retainers
Ricochets (Dynamic objects of fear) Goal is *find cover*	Plan is *make for cover* Uses *doorways, walls, alleyway, etc.* Work is *navigation skills*	Activity is *take cover* (Local) Reward is *time to think, plan, etc.*
Movement of opponent(s) (Dynamic object(s) of fear and desire – your opponent can fight back) Goal is *find cover*	Plan is *make for cover* Uses *doorways, walls, alleyway, etc* Work is *navigation skills*	Activity is *take cover* (Local) Reward is *time to think, plan, etc.*
Movement of opponent(s) (Dynamic object(s) of fear and desire – your opponent can fight back) Goal is *frag opponent*	Plan is *take opponent by surprise* Uses *guns and ammo and maybe cover.* Work is weapons skills and navigation etc.	Activity is *firefight* (dynamic, peripatetic) Reward is *fun + increase frag count*

The possible relationships between attractors gives us differing structuring mechanisms that can form the basis of narrative potential. At any one time we may have a choice between a number of different attractors or a choice of responses to the same attractor. These equate to Janet Murray's choice points [2]. Groups of attractors and their associated rewards may form an identifiable task or action and equate with the mini-missions of the computer game world. Particular attractors may instigate challenge points, which are particular goals that have to be attained to make further progress in the VE possible.

Rather than pursue these mechanisms and their relationship with narrative potential and virtual storytelling in the abstract we will proceed to the next section where they will be illustrated with reference to a computer game that would appear to exemplify the very possibilities for virtual storytelling we are looking for.

6 The Perceptual Opportunities of Shenmue

Shenmue is a computer game, a role-playing game, and is interesting because it appears to challenge the notion that computer games cannot tell stories. In this section we will apply POs to Shenmue and come to some conclusions concerning the basis on which virtual storytelling can become a reality (again, the play on words is deliberate).

Shenmue is a quest in which we, the player, direct the principle protagonist, Ryo Hazuki[2] in his endeavors to find his father's murderers. Shenmue is a vast, interactive 3D virtual environment in which the player has to search out clues which will lead him/her to Ryo's father's killer. Despite the extensive reliance on agency Shenmue is has many of the characteristics of narrative. We have a plot, based around the quest, a genre, the detective story, we have characterisation, and we have a beautiful evocation of not only the architecture of neighborhoods but also of the extensive social relationships that are the true heart of those neighborhoods. In Shenmue we have all these characteristics of narrative existing side by side with agency. Why should this be so despite the oft-cited intuition that games and narrative are mutually exclusive?

In the previous section we discussed the relationship between the pleasure of agency and the expression in the VE of attractors, connectors and rewards. If we look at Shemue for evidence of these we immediately come up against the problem of their sheer number and density. However, we can categorise all these into a few general types, which are:

- Examining and purchasing inanimate objects
- Interacting with active objects such as doors
- Talking to people
- Quick timer events
- Free battles
- Playing arcade games

The majority of work in Shenmue is concerned with the first four. Despite the slight differences in the means of interaction these four share a very interesting characteristic - they all reward users' exercising agency with a pre-defined sequence

[2] As with Shenmue, Ryo Hazuki is a trademark of Saga of America Inc.

of actions which effectively temporarily removes users' ability to exercise agency. For instance, in the basic act of opening a door we have the following sequence of events:

1. We perceive an attractor, the door, within the field of view
2. We approach the door
3. As we come within close proximity to the door an icon representing the 'red A' button on the controller appears close to or over the door
4. We press the actual 'red A' button on the hand controller
5. The game engine rewards the users' action with a pre-defined sequence of Ryo positioning himself in front of the door, turning the door handle and opening the door, walking through the door and then closing it behind him. We can only sit and think while this sequence runs to conclusion.
6. The game engine then loads the files, which represent whatever is on the other side of the door.

Why is this so interesting? Well, in most computer games we would trigger a sensor or touch a switch and the door would open and we would walk through. But all this would be under our own volition and if we got in the way of the door we might accidentally stop it opening properly and perhaps be injured in the process. In Shenmue we loose control of the details of the act. Agency is rewarded by removal of agency. But this is exactly the interface of agency and narrative at the basic level of the units, lexia, of the perceptual opportunities of the game. Agency is rewarded by a narrative fragment.

Apart from free battle interludes and playing arcade games, the result of exercising agency is always a pre-defined, sometimes a pre-rendered, sequence. If we are talking to people this will mean a question from Ryo followed by some sort of response, not necessarily helpful or polite, from the person he has spoken to. The conversation can often be continued by another press of the 'red A' that will result in another question and response.

Multiple acts of agency are rewarded by the build up of more and more of these fragments all of which in their own way contribute to the narrative potential of the game. Unlike typical *shoot-em-ups* and *sneak-em-ups* narrative components are not simply used to frame whole game levels or major subsections of levels. Narrative components are integrated into the game at the level of agency.

One of the consequences of this interplay of agency rewarded by narrative fragments is that the game can use extended *cut scenes* to introduce more substantial narrative material without interrupting the flow of the game. We are simply getting a bigger reward. Cut scenes can also be introduced for other reasons than agency. For instance, the fall of night is indicated by a cut scene of the night skyline of the particular district we are in. The playing of the cut scene is triggered by the time of day - Shenmue time, which runs a lot faster than real time - yet it is not a shock but a pleasant surprise made possible by the basic nature of attractors and rewards which pervade the game.

One of the main reasons for the rich levels of connotation of Shenmue is because conversations and therefore language play a central role in the information space of the game. Further, Shenmue does not have distinct levels but offers a continuous flow of interaction limited only by the storage capacity of the three CD-ROMs on which it is delivered.

Shenmue also makes use of interactive variations on the cut scene idea. Quick Timer Events, for instance, occur in certain situations and require the player to recognise an icon as representing a particular controller button flashed on screen and then press the actual button within a fraction of a second. We usually get several goes at this until we 'get it right'. Examining and picking up and buying objects also works is a similar way as interactive pre-defined sequences.

There are other interesting points to note about Shenmue. It is quite unusual for the physical interface, the controller, to be represented within the game itself. It would normally remind players that this world is mediated and that button presses are analogues for walking and running, for instance. We do not always know the pre-defined sequence we are to get. If the door is locked we might get Ryo's thoughts, a request to go away from the other side of the door and so on. This uncertainty of the outcome of exercising agency is used to great dramatic effect in Shenmue. The encroachment of the 'red A' into the game world is also used to highlight possibilities for agency, which are not obvious from the game logic the player has so far encountered.

7 POs, Narrative Potential, and Virtual Storytelling

POs offer a view of the basic components or lexia of VEs in terms of agency. The organisation of these into a perceptual map allows us to consider their configuration in terms of larger structures - such as routes, choice points, challenge points and retainers - that represent the narrative potential of a VE. Narrative potential can be seen as both the degree to which such structure can accumulate to form meaningful experience and the degree to which content preserves its meaning over the course of the narrative rather than being overwhelmed by the pleasure of agency.

We can now identify a number of similarities between traditional narrative and virtual storytelling - at least in the context of Shenmue. Both have:

- Extensive characterisation – at the heart of Shenmue are the diverse range of distinctive characters, including Ryo, who we have to get to know and understand.
- Levels of meaning based on connotations not directly expressed in the text or game – we come to see the neighborhoods of Shenmue as social spaces and not geometrical, for instance.
- A confusion of the consequential and the consecutive – because narrative is integrated at the level of agency we don't immediately know what is important and what is not.

This is not to say that virtual storytelling is just narrative on computers for it is not. There are major differences:

- Barthe asserts that when a traditional narrative reaches a major choice point between alternative actions it always makes the choice that ensures it continued survival. This is clearly not the case with virtual storytelling because of agency.
- In traditional narrative forms agency is reduced to the decision to read on, view on, listen on, and so on, or not whereas virtual Storytelling requires the active expression of agency.
- This has the result that in virtual storytelling, challenge points are genuine challenges, which cannot be resolved by reading or watching on. As the player I have to solve the problem before I can proceed.

- One of the characters takes on a particular significance because it the one associated with the playable character and therefore indirectly with us the player. This is interesting because the playable character is not me with some cyborg-like exosuit, my avatar and its accoutrements, strapped over my existing self but rather an external character who I can emphasise with and control to a certain extent. But the playable character is most definitely not me, however much I empathise with him.
- Virtual storytelling, at least in the case of Shenmue, has both a relative narrative time - a consequential ordering of events - and a continuous and pervasive *real* time which, quite literally, tick mercilessly away at the bottom right hand corner of the screen. Traditional narrative has only narrative time.

8 Conclusions

In this paper we investigate the notion of virtual storytelling seen as narrative potential, which is deemed to be the reconciliation between agency and traditional narrative forms. In order to do this we have applied the perceptual opportunities model of VE content to the popular computer game Shenmue. This allowed us to characterise the basic nature of agency in Shenmue. We observed that agency in Shenmue almost always rewards action with a narrative fragment, a pre-defined or pre-rendered sequence, which has the effect of removing agency temporarily. By structuring the basic units of agency in this way the player learns to take pleasure in the accumulation of clues and information towards the resolution of the quest before being returned to a situation of agency in order to proceed.

Of course, we have only one computer game but we can begin to draw some tentative conclusions as to the basis on which virtual storytelling from a first person point of view is at least possible:

- Agency and narrative must be integrated at the level of basic units
- Cut scenes, even long ones, then become just bigger rewards integrated into the game itself
- An integrated flow of development (no levels as in traditional agency focused games)
- The extensive use of language in the form of conversational fragments, tones of voice and body language greatly increases characterisation and connotation and thus increases the richness and level of meaning associated with more traditional narrative forms.

Shenmue is a quest. Could we envision a virtual storytelling that was a psychological thriller? How about a virtual storytelling in the manner Ben Okri's mystical streams of consciousness? Can we make an effective equivalent of Jack Kerouac's 'On The Road' where the quest is about self realisation rather than the specific, measurable goals of the detective story? It is the belief of the author that such virtual storytellings are possible but that the true integration of agency and narrative must be a subtle affair. Narrative potential can be thought of as the study of the ecology of narrative - the study of the conditions under which agency and narrative may thrive.

References

1. Jull, J.,: A Clash Between Game and Narrative (Why computer games do not tell good stories and why this is not a problem, Conference on Digital Arts and Culture, Bergen, (1998)
2. Murray, J.,: Hamlet on the Holodeck: the future of narrative in cyberspace, Free Press, (1995)
3. Davenport, G., Agamanolis, S., Barry, B., Bradley, B., and Brooks, K.,: Synergistic Storyscapes and Constructionist Cinematic Sharing, IBM Systems Journal, http://www.research.ibm.com/journal/sj/393/part1/davenport.html, (2000)
4. Barthes, R,.: Introduction to the Structural Analysis of Narratives in: The Semiotic Challenge, translated by Richard Howard, Basil Blackwell, (1988)
5. Barthes, R.,: S/Z. Noonday, (1991)
6. Aarseth, Espen J.,: Cybertext: Perspectives on Ergodic Literature. John Hopkins University Press, (1997)
7. Fencott, C., Shaffiulah, M., van Schiak, P., & Ling, J.,: The Effects of Attractors and the Pictorial Content of Retainers on User Behaviour in Virtual Environments, submitted for publication, (2001
8. Isdale, J., Fencott, C., Heim, M., & Daly, L.,: Content Design for Virtual Environments, to appear in: The Virtual Environment Handbook, ed)(K. M. Stanney, Lawrence Erlbaum Associates Inc, (2001)
9. Church, D,.: Formal Abstract Design Tools, Games Developer Magazine, August,(1999)
10. Fencott, C.,: Content and Creativity in Virtual Environment Design, in proceedings of Virtual Systems and Multimedia '99, University of Abertay Dundee, Dundee, Scotland, (1999)

Adaptive Narrative: How Autonomous Agents, Hollywood, and Multiprocessing Operating Systems Can Live Happily Ever After

Jay Douglas and Jonathan Gratch

University of Southern California
Institute for Creative Technologies
Marina del Rey, California 90292
{jdouglas, gratch}@ict.usc.edu

Abstract. Creating dramatic narratives for real-time virtual reality environments is complicated by the lack of temporal distance between the occurrence of an event and its telling in the narrative. This paper describes the application of a multiprocessing operating system architecture to the creation of adaptive narratives, narratives that use autonomous actors or agents to create real-time dramatic experiences for human interactors. We also introduce the notion of dramatic acts and dramatic functions and indicate their use in constructing this real-time drama.

1 Introduction

```
EXT - BOSNIAN VILLAGE STREET - DAY

A young lieutenant is on his way to a rendezvous with the rest
of his platoon near the village square. His RADIO crackles out
an assignment.

                        RADIO VOICE
               We need you here at the armory
               as soon as possible.

But the lieutenant, still a few kilometers away, is preoccupied.
We SEE a traffic accident involving one of the lieutenant's
humvees and two local CIVILIANS. One, a YOUNG BOY, is seriously
injured and hovering near death. The second, his MOTHER, is
unharmed, but in shock and hysterical. A menacing CROWD gathers.
A CAMERAMAN for an international cable channel materializes,
shooting tape for the evening news.
```

This isn't a snippet of a Hollywood movie script. It's part of an interactive story based on real-life experiences of troops assigned to peace-keeping missions in the former Yugoslavia. In this tale, a lieutenant faces several critical decisions. The

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 100–109, 2001.

platoon at the armory is reporting an increasingly hostile crowd and requests immediate aid. The boy needs medical attention, which could require establishing a landing zone prior to a helicopter evacuation. The accident area must be kept secure, but excessive force or a cultural faux pas could be construed as a cover-up with major political consequences. The lieutenant's orders prohibit the use of weapons except in the face of an immediate threat to life or property. Unlike the movies, though, this cast, with the exception of the lieutenant, consists entirely of computer-generated characters, several of which are autonomous and cognitively aware. Instead of a Balkan village, the action takes place in a virtual reality theater with a 150-degree screen, 3 digital Barco projectors and an immersive, surround sound audio system equipped with 10 speakers and two sub-woofers (a 10.2 arrangement as compared with the typical 5.1 home theater system). The agents can interact with the lieutenant through a limited natural language system, and the mother agent responds to changes in her environment in a limited way. In spite of all this technology, we still cannot guarantee our storytelling environment will deliver engaging, dramatic content. This paper presents our work-in-progress on content production for the Mission Rehearsal Exercise (MRE) [1], one of the major research efforts underway at the Institute for Creative Technologies (ICT).

What we describe here is a multiprocessing operating system-like architecture for generating story world events unknown to the interactor, and the notion of dramatic functions, a method for gathering these events into dramatic moments. These low-level tools allow a human in the storytelling loop to create dramatic content in a real-time environment.

2 Motivation

Regardless of the medium, literary theorists describe the creation of narrative content as a three-step process: selection, ordering, and generating.[1] Out of all possible occurrences in the story world, some set must be selected for telling. Next, these occurrences must be ordered. There is, after all, no requirement that a story unfold chronologically. With apologies to Aristotle, we can begin at the end, then jump back to the beginning. If we choose to organize our narrative in this way then we require a crucial condition be met: the narrator must have a temporal distance from the events, where temporal distance means the events in the narrative occurred at some time other than the time of their telling.

In traditional storytelling this is no problem, for the telling of a tale comes after the occurrence of the events comprising it. If all occurrences unfold in real time, the processes of ordering and selecting are governed more by physical, rather than narrative, concerns. Our ability to create mystery, suspense, humor, and empathy are compromised.

Rather than abandon these powerful literary devices, our goal is adapting these techniques to the context of a real time environment. To do this, we need

[1] The steps are taken from Bordwell [2], substituting "generating" for his term "rendering" to avoid confusion with graphics terminology.

to maintain a separation between the time of events and the time the interactor learns about them. Once an event becomes "public," we forfeit the chance to foreshadow it, and recognizing foreshadowing opportunities is complicated by the interactor's freedom of choices. One apparent solution is providing the interactor with physically partitioned spaces into which he or she can move and ask "What happened here?" Events in these spaces would be known to us and temporally distant from the interactor so we could construct our dramatic moments. Such an approach leads to narrative consistency problems. Very quickly, we can wind up with a collection of moments each inconsistent with those of other spaces. What we suggest is creating "potential" foreshadowing opportunities to serve as fodder for our narrative content.

3 An Adaptive Narrative Architecture

A viable source for such foreshadowing opportunities presented itself unexpectedly, as side effects of a series of Wizard of Oz (WOZ) experiments.[2] A number of autonomous agents were replaced by human actors, and scenarios were played out under the invisible control of a (human) wizard. An agent command interface and two-way radios closed the behavior loop, giving the wizard control over agents and actors, as well as over the timing of interactions between them. The unexpected drama we encountered encouraged us to build an infrastructure for playing out multiple scenarios in parallel, under the control of a software *narrative agent* capable of cranking through the dramatic functions and turning events into dramatic experiences.

In life, unlike in books or films, the world goes on outside the pages we read or the images accessible to us on the screen. In books and film, moreover, readers and viewers only know what the author/director wants them to know. Not so in life (or adaptive narratives). If the interactor hears a noise behind a door, he or she should have the option of discovering the source. This may mean opening the door, asking another character, or seeing "through the door" via a surveillance camera. While the reconstruction of life may tax our abilities and our patience, our WOZ experiments pointed the way to a more user-friendly computer science model: the multi-processing operating system.

In UNIX-flavored systems, the user may have one process running in the foreground, but many others operating in the background. Similar effects were recognized in our WOZ experiments. We always had a foreground narrative, one involving the lieutenant, while other narratives, background processes in effect, played out somewhat unnoticed and asynchronously "offstage." These background narratives unwound according to their own scripts, and even though their actions were not the focus of the lieutenant's attention, their unfolding generated events the wizard used to increase or decrease the lieutenant's stress level.

[2] Although these experiments were performed to collect data for dialogue systems research, the results that intrigued us were those similar to Kelso [3].

Our developing system model relies on the abilities of autonomous agents to carry on their "lives" outside of the main focus and control of a central authority. By allowing these agents to execute their own scripts somewhat out of sight, the narrative agent accumulates invisible (to the interactor) events to support dramatic effects. Our background narratives run independently of each other, eliminating timing and contention problems. In our current design, the frenzy of the crowd at the accident scene, the situation at the armory, the attempts of a TV news cameraman to interject himself into the situation, and the status of men and equipment attached to the base commander all vary at their own speed, based on parameters established by the narrative agent. Thus, the cameraman agent might accede to a soldier's order back away from a shot if the cameraman's desperation factor is low (he's got his most important shots), or hold his ground if getting the shot means earning enough for his baby son to eat that night. While the lieutenant can "swap" foreground and background narratives, in same way as the fg and bg console commands can swap UNIX foreground and background processes,[3] a background narrative can always create an "interrupt," demanding attention from the lieutenant. For example, a background character can initiate a conversation with, or send a radio message to, the lieutenant, immediately bringing itself to the fore. Perhaps most importantly for training and education, modifying agent attitudes and the speeds of background narratives means each retelling of the MRE tale opens up new interpretations, each still causally linked to the interactor's behavior, each with its own feel and intensity, and each created without additional scripting or programming.

3.1 Choosing Background Processes

While any background narratives might suffice, at least in theory, we want to constrain them so the events they generate lend themselves to drama in the specific narrative we are working on. One way to accomplish this is to choose background narratives congruent with the interactor's goals. In the MRE, the interactor wants to: (a) evacuate the boy, (b) maintain local security, (c) fulfill his responsibilities relative to the platoon at the armory, and (d) perform in a manner consistent with good public and press relations. Starting with these goals, our scenario writers created five background narratives: (a) a threatening situation at the helicopter landing zone caused by any number of sources, from belligerent crowds leaving the armory to engine failure; (b) a mob scene at the accident site caused by provocateurs inciting an initally curious crowd; (c) an increasingly dangerous situation at the armory, where crowds are growing in size and their demeanor is growing more threatening; (d) an aggressive TV news cameraman who insists attempts to restrain him are actions preventing him from reporting the true story of military arrogance; and, (e) a deteriorating situation at base command, where demands on men and equipment may mean a shortage of relief troops, ground vehicles, and helicopters. On their own, the

[3] Swapping foreground and background occurs when the lieutenant interacts with a character in a background narrative.

background narratives are independent of each other; however, their common focal point is the interactor. He may alter the status of one narrative based on events occurring in another. All the narratives, however, affect the interactor's ability to meet his goals. They provide the fodder for what is typically known as the drama's "second act," the part where the protagonist embarks on a certain course, passes the point of no return, and finds his way strewn with obstacles.

For the narrative agent, however, the great advantage is the interactor's relative ignorance of events occurring offstage. Unless the interactor checks, he doesn't know the state of affairs at the armory. The narrative agent does, however, so if the interactor issues a radio call to the armory, the results are liable to come back garbled. The snatches of understandable transmissions may yield the wrong impression of the scene. Or, we might find the cameraman insistent on getting a shot based on something whispered to him by the boy's mother. A high ambient noise level, stress, misunderstandings, all are at the narrative agent's disposal for presenting a narrative to the interactor of the agent's own making.

We still, however, need guidance in selecting and ordering these events. For this we introduce the notion of dramatic functions.

4 Dramatic Functions

An old writer's adage says that if you plan to shoot a character in the third act of your play the audience had better see the gun in the first act. It's a reminder that unmotivated actions appear to come from "out of nowhere" in drama. In the same vein, coincidences, obstacles, misperceptions, misunderstandings and other storytelling tools sometimes test the bounds of credulity even while creating engaging narrative experiences. We perform acts and create situations in narratives that might be judged exaggerated in real life. Gerrig [4] discusses one theory of why we, as readers, viewers, or interactors, easily accept this distortion of reality and why it does not interfere with our enjoyment and involvement. Thus, in drama we find two types of acts: acts that occur as they might under real circumstances, and *dramatic acts*, which are manipulations necessary to create emotional responses. In our work we employ the notion of *dramatic functions* to construct dramatic acts. Representing drama as a set of base dramatic functions is one of the contributions of this implementation to virtual storytelling.

4.1 A Functional Approach

When one talks about describing functional elements in narratives the work of Vladmir Propp [5] springs to mind. Working with the Russian folk tale, he identified 31 actions played out by specific character types. The actions and characters generated hundreds of tales by simply changing settings or personalities. Propp's research also discovered these actions, if they appeared in a tale, appeared in strict order. Number five always occurred after any lower-numbered functions and before any higher-numbered ones. Because of this rigid structure, Propp's functions only generate a specific narrative form. Their greatest contribution to virtual storytelling, however, is the notion that narratives can be described in functional form.

Szilas [6] carried Propp's work several steps further by developing a set of generalized functions for constructing narrative content. The general direction of his research informs our own work in constructing our narrative content. Szilas's functions are broadly applicable to content behind the narrative, such as descriptions and chronicles. Our search is for something more middle of the road, an approach somewhere between the restrictiveness of Propp and the generality of Szilas. In addition, we want a system that allows us to reason about temporal relations as well as propositions and beliefs. Towards that end we asked the question: what makes drama different from real life?

4.2 Dramatic Function Notation

One of the characteristics of dramatic acts, and hence dramatic functions, is their time dependency. The villain and his machete do not materialize until it appears the occupants of the haunted house need only open the front door and escape. James Bond doesn't disarm the bomb until there is no time left on the timer. Not only do we need the classical notion of events occurring before or after one another, we must reason about how long the separation between event and knowledge of the event should be and deal with events that occur over time rather than instantaneously. Thus, we require a logic that not only admits time intervals, but one robust enough to describe such commonplace relationships as during, meets, shorter, and longer. In order to reason about events in their temporal context, we represent our dramatic functions along lines outlined by Allen [7].[4]

In his temporal logic, Allen describes three primitive functions in their logic: $\mathrm{OCCUR}(e, t)$, $\mathrm{OCCURRING}(r, t)$, and $\mathrm{HOLDS}(p, t)$.[5] OCCUR is true if event e occurs strictly over interval t (that is, for any t_i less than or equal to t, $\mathrm{OCCUR}(e, t_i)$ is false). $\mathrm{OCCURRING}(r, t)$ is true if process r takes place during interval t; however, it is not necessary for r to happen at every subinterval of t. A process is distinguished from an event because we can count the number of times an event occurs. $\mathrm{HOLDS}(p, t)$ is true if property p is true over interval t (and all subintervals). We add a new primitive to this collection, $\mathrm{ASSERT}(a, b, p, t)$. For this event, during interval t, a asserts to b that proposition p is true. If ASSERT is true then we can conclude that a succeeded in asserting p to b and the act took place during interval t. Note that the function makes no claim about whether b believes a.

[4] Allen rigorously defines arithmetic on time intervals, as well as concepts such as before, after, during, and meets in his paper. For clarity, we omit his axioms and appeal here to intuitive definitions.

[5] We use lower-case letters to denote variables and uppercase letters to denote bindings.

Finally, we define a new function BELIEVE(i, p, t), which is true if a human or agent interactor i believes that proposition p holds over interval t. Modeling belief is a non-trivial undertaking, so in our work we rely on a fairly restrictive definition: BELIEVE(i, p, t) is true during interval t if p is not contradicted by any information presented in the domain during interval t and ASSERT(a,i,p,t_p)is true, where t_p precedes and meets t.

We do not deny the definition lacks a certain sophistication, especially the first clause, which presupposes a deficit of domain knowledge on i's part. The danger of such an assumption is obvious in general. We find it acceptable in the present case, because a fundamental motivation behind the MRE is that the interactor knows very little about the non-military elements of the story world. If we restrict the use of BELIEVE to propositions ranging over this domain, the definition becomes manageable, if not quite reasonable.

In the next section we give the general definitions for two dramatic functions, reversal and snare. Later in this paper we will apply these functions to the MRE in a concrete example.

4.3 Reversal Function

In a reversal (often called a reversal of fortune), the interactor sees the attainment of a goal snatched away from her—usually moments before success is at hand. Thus, for all events E, where E is a goal of the interactor, such that P is the set of preconditions, and $|t_s|$ is the length of time that event E will take, iff

$$\text{HOLDS(P,}t\text{)} \wedge |t| \geq |t_s| \rightarrow \exists t_s\text{:OCCUR(E,}t_s\text{), where } t_s \text{ is a subinterval of } t.$$

In words, once the preconditions of E are satisfied, and remain satisfied during the time it takes for E to occur, E will occur. Thus, the interactor should *expect* a successful outcome, especially if there is no perceived threat rendering HOLDS(P,t) false.

The narrative agent's role is reversing this expectation. While the interactor expects E, the narrative agent plans

$$\text{HOLDS(P,}t_1\text{)} \rightarrow \text{HOLDS(}\neg\text{P,}t_2\text{)}$$

where $|t_1| + |t_2| = |t|$, $|t_1| < |t_2|$, and t_1 precedes t_2 and t_1 meets (is adjacent to) t_2. Since P becomes false during the interval in which E requires P to be true, the goal is thwarted.

4.4 Snare Function

A snare is a misrepresentation, a deception, usually one that deepens a mystery.[6] In our application, a snare represents an attempt by the narrative generator to lead the interactor astray (and thwart one of his goals) by deliberately presenting the world as it is not.

[6] See Barthes [8] for a discussion of snare and other function types.

Let P be the set of preconditions for E, where E is a goal of the interactor, I. By the reasoning above, the interactor expects E to occur because

$$\forall i\text{:BELIEVE}(\mathrm{I},P_i,t_s)$$

Let us also define P' such that

$$\forall i \neq j\text{:BELIEVE}(\mathrm{I},P_j,t_s) \wedge (\neg P_j \wedge \text{BELIEVE}(\mathrm{I},P_j,t_s))$$

In the snare, the narrative agent's role is to construct a P' based on P, such that the interactor believes Pj and ultimately expects E, whereas the truth is P' (and therefore $\neg E$).

5 Concrete Examples

To see how dramatic functions and background processes work together let's consider two possible sequences in the MRE. In the first, the interactor orders a helicopter evacuation of the boy. Let P be the conjunction of the four conditions: landing zone (LZ) is empty; LZ is marked with green smoke; LZ surrounded by one squad of troops; and, helicopter is over the LZ. When all these conditions are met, the helicopter can land.

The narrative agent must reverse E, the event "helicopter lands" at the last possible moment. The agent has, for this narrative, the following domain information: a crowd of 50 people are marching from the armory and are only a few blocks from the LZ; helicopters are complex machines and can develop problems requiring them to return to base and the base is not near the LZ. The narrative agent must search for a plan that results in $\neg P$. Which one to choose is a matter of style and dramatic familiarity. Mataes and Stern [9] suggest that when creating conflict one should choose situations referencing past occurrences. The story agent might check its history to see if, for example, the interactor was warned about overextending his troops (recommendation from platoon sergeant), or if the interactor was warned by the armory platoon leader that a crowd was heading towards the accident scene (background narrative), or if the base reported it was having trouble keeping its helicopters mechanically sound (background narrative). Since E is associated with an interval over which it occurs, the narrative agent can reason about not only how to create $\neg P$, but when to create it as well.

A common Hollywood use of the snare is the false ally, in which an antagonist presents herself to the protagonist as a friend while secretly working to foil the protagonist's plans. As an example of an MRE snare, consider what happens when a crowd of villagers forms at the accident site. Since neither side speaks the other's language, the crowd can only judge the soldiers' intentions towards the young victim by observing their actions; and, the soldiers can only judge the crowd's intentions by interpreting body language and tone of voice. Here is a situation ripe for misunderstandings. A restless crowd, unfamiliar with military medical techniques, on one side, nervous soldiers easily capable of misinterpreting emphatic, but harmless, gestures on the other. Certainly, it is in the interactor's best interests to keep the crowd calm.

Let E be the goal "trusted person tells crowd boy is getting good care," which we will denote by BELIEVE(C, a, B, t), where C is the crowd, a is an agent (possibly the interactor), B is the proposition "boy is getting good care," and t is some time interval over which the belief holds.

A priest, speaking broken English, materializes from the crowd and offers his services (background narrative). He will inspect the boy and report back to the crowd on the aid being administered. What the interactor does not know is the priest is a provocateur and will incite the crowd no matter how attentive the soldiers are to the boy's needs. The narrative agent expects the interactor will trust the priest (domain knowledge) and therefore will expect E is achieved over a long interval, t, once the priest talks to the crowd. However, the priest's words will be interpreted as inflammatory by the agents controlling the crowd's behavior. Sometime in t the crowd's fury will boil over (as determined by the narrative agent), hopefully surprising and distressing the interactor.

The narrative agent functions only if it recognizes the interactor's current goals, and this recognition represents another open issue. In a general solution, a plan model would provide feedback to the narrative agent about the interactor's intentions. We have no such mechanism, but we do have the structured domain of a military operation. In the MRE, the interactor's goals are typically expressed as orders, as when the interactor orders a helicopter evacuation. Recognition of these orders is necessary for other parts of the MRE, and the narrative agent can piggyback on these software modules, grabbing the orders as necessary and turning them into goals to be manipulated.

6 Future Work

What we've outlined here is only part of the story. The three-step narrating process as a model for a multi-processor-like storytelling environment, the notion of potential foreshadowing, and the encapsulation of primitive elements of drama into dramatic functions are promising tools; however, we believe a fully autonomous narrative agent is within reach of the state of the art. Notwithstanding our optimism, the future finds us with major obstacles to overcome. So far, we have not considered how the narrative agent combines the use of dramatic functions into a cohesive narrative. Mataes and Stern [9] provide a clue, but for complete generality, an agent will need to make far more subtle decisions, such as which dramatic function to choose for a particular effect, when to inject drama and when to allow the narrative room to "breathe," how far ahead to look when planning dramatic content, and how to recover when the interactor upsets the narrative agent's plan. Right now, a human still needs to interpret the interactor's goals in order to thwart them. The general recognition problem is still an open issue, and will most likely entail a model for recognizing narratives being constructed in the interactor's mind [4], [10], [2] combined with a mechanism for sharing these narratives, along the lines described by Young [11].

Despite these challenges, our research inches us closer to narrratives with more dramatic content than currently available, and to narratives that vary considerably in the "retelling," without the need for reprogramming or re-scripting. While there remains much work to be done, the combination of knowledge from both computer science and Hollywood offers exciting possibilities for the future of virtual storytelling.

Acknowledgements. The authors wish to thank David Traum and Michael van Lent for their assistance in clarifying and formalizing many of the concepts in this paper, and for their patience in reviewing and commenting on the manuscript.

References

1. Swartout, W., et al.: Toward the holodeck: Integrating graphics, sound, character and story. In: Proceedings of the Fifth International Conference on Autonomous Agents, ACM Press (2001)
2. Bordwell, D.: Narration in the Fiction Film. U of Wisconsin Press, Madison, Wisconsin (1985)
3. Kelso, M.T., et al.: Dramatic presence. Technical Report CMU-CS-92-195, Carnegie Mellon University (1992)
4. Gerrig, R.: Experiencing Narrative Worlds: On the Psychological Activities of Reading. Yale UP, New Haven (1993)
5. Propp, V.: The Morphology of the Folktale. 2nd edn. U of Texas Press (1988) Trans. Laurence Scott.
6. Szilas, N.: Interactive drama on computer: Beyond linear narrative. In Mateas, M., Sengers, P., eds.: Narrative Intelligence. AAAI Fall Symposium, Menlo Park, California, AAAI (1999)
7. Allen, J.F.: Towards a General Theory of Action and Time. In: Readings in Planning. Morgan Kaufmann (1990)
8. Barthes, R.: S/Z. Hill and Wang, New York (1974) Trans. Richard Miller.
9. Mateas, M., Stern, A.: Towards integrating plot and character for interactive drama. In Dautenhahn, K., ed.: Socially Intelligent Agents: The Human in the Loop. AAAI Fall Symposium, Menlo Park, California, AAAI (2000)
10. Branigan, E.: Narrative Comprehension and Film. Routledge, London and New York (1992)
11. Young, R.M.: Using plan reasoning in the generation of plan descriptions. In: AAAI/IAAI, Vol. 2. (1996)
12. Chatman, S.: Story and Discourse: Narrative Structure in Fiction and Film. Cornell UP, Ithaca and London (1980)
13. Genette, G.: Narrative Discourse: An Essay in Method. Cornell UP, Ithaca (1980) Trans. Jane E. Lewin.

Learning in Character: Building Autonomous Animated Characters That Learn What They Ought to Learn

Bruce M. Blumberg

The Media Lab, MIT
77 Massachusetts Ave., N18-5FL Cambridge Ma. 02139
bruce@media.mit.edu

Abstract. We suggest that the ability to learn from experience and alter its observable behavior accordingly is a fundamental capability of compelling autonomous animated characters. We highlight important lessons from animal learning and training, machine learning, and from the incorporation of learning into digital pets such as AIBO and Dogz. We then briefly present our approach, informed by the lessons above, toward building characters that learn. Finally, we discuss a number of installations we have built that feature characters that learn what they ought to learn.

1 Introduction

If presented with an autonomous virtual character such as a virtual dog, people expect the character to be able to learn the kinds of things a real dog can, and alter its behavior accordingly. The reason, simply put, is two-fold: first, people expect a level of common-sense from animals such as dogs, and second, the very way we understand any animate system is by assuming that it will behave so as to “get the good” and “avoid the bad” given its desires, repertoire of actions, and beliefs about how the world works. In addition, we expect that the creature will revise its beliefs based on experience. Indeed, the ability to learn from experience is one measure of what people often label as intelligence, i.e. more intelligent creatures are better able to learn than less intelligent creatures.[1] When a character doesn’t learn from experience, we are left wondering “is it stupid, or is it simply broken.”

Much of the work of the Synthetic Characters Group at the Media Lab of MIT is devoted to understanding how to build autonomous animated characters that can learn what they ought to be able to learn. We take our fundamental inspiration from animal learning and training. Our belief is by paying close attention to how animals learn, and successful techniques by which they are trained, we can not only improve on existing models for machine learning, but also develop robust techniques for real-time learning in autonomous animated characters.

In this paper we begin by presenting the case that an autonomous character’s ability to modify its beliefs based on experience is a fundamental requirement for any

[1] Note, that while we speak of “learning from experience,” our measure of this is the extent to which the creature alters its actions in a manner that we believe makes sense given our understanding of its goals and its experience.

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 113–126, 2001.

character whose actions are intended to be a direct and believable consequence of its desires and beliefs. We then present lessons from experience with digital pets, machine learning and finally from animal learning and training that we believe are useful for guiding efforts to build autonomous characters who are perceived as learning what they ought to learn. This is followed by a brief summary of our approach, and finally a discussion of our experiences to date building characters that learn.

2 Why Characters Need to Learn: Learning and the Intentional Stance

Classics such as The Illusion of Life [16] explain the art of revealing a character's inner thoughts—its beliefs and desires—through motion, sound, form, color and staging. While the "Illusion of Life" makes it clear what one must do if one wants to bring a character to life, it does not address the question of why these techniques work. A concise explanation can be found in the work of the philosopher Daniel Dennett [4]. Dennett argues that the "Intentional Stance" is the fundamental strategy we use to predict and explain the actions of animate systems including people, animals, and animated characters. The Intentional Stance is simple. First, one decides what goals or desires the character ought to have and what set of actions it can perform. Then, one decides what set of beliefs the character ought to have about the effect of its actions on the world and ultimately on its desires. Finally, one assumes that it will always act in a commonsensical way [given its character] so as to satisfy those desires given its beliefs. Seen in this way, we use the Intentional Stance to predict a character's actions based on our knowledge of its presumed desires and beliefs. Conversely, we use it to infer a character's desires and beliefs based on our assumption that its motion and the quality of that motion is a direct consequence of its those desires and beliefs. In the context of Dennett's work, one sees that the techniques put forth in the Illusion of Life are essentially a recipe for making it easy for the viewer to take the Intentional Stance relative to a character. Because of their central importance it is worth spending a bit of time outlining what we mean by desires, beliefs and actions.

2.1 Desires

Desires arise from internal needs: hunger, thirst, sex, need for revenge etc. Every great character has character-specific desires that they attempt to satisfy in character-specific ways. Narratives often revolve around how a character ultimately satisfies its desires by learning how to overcome obstacles standing in its way, or how the fundamental nature of a character's desires change as a result of learning.

2.2 Beliefs

Just as every character is endowed with character-specific desires that help define what it means to be that character, so too are they endowed with character-specific beliefs about the world and their place in it. Beliefs are what connect desires and actions. Essentially beliefs are the shorthand for the mechanism that, given the state of the world and the state of the character, decides what actions should be taken so as to satisfy the underlying drives in a commonsensical manner. Thus, beliefs should reflect:

1. Perceptual input, e.g. "If I see a stream, then I believe I will find water there",
2. Emotional input e.g. "Because I am afraid of snakes, I should not pick up that snake", and
3. Experience e.g. " The last time I was in this field I saw a snake, so I am sure he is there today".

We assume characters perceive those things that "it makes sense" for them to perceive given their character and the state of the world, and that their beliefs are adjusted accordingly. That is, we assume they perceive what *we* would perceive in a similar situation, and that their beliefs *are altered as a result just as ours would be in the same situation.*

Just as the discovery of desires is the grist for many stories, so too is the resolution of a mismatch between a character's beliefs about its world and the reality of that world. These stories only work because of our use of the intentional stance as a means of understanding the character. They also rely on the viewer being let in on the secret early in the game and so that they are in a position to understand (and laugh at) the seemingly inappropriate actions of the character. And ultimately, they only work because the character learns.

2.3 Actions

Beliefs and desires are revealed through one thing alone, and that is the character's actions, or more precisely through the character's actions as revealed by the staging and cinematography.

As Rose points out [13], animation is all about "verbs and adverbs" and hardly ever just about "verbs". A well-animated character never just walks, rather there is always an adverb that can be easily associated with the walk, e.g. "angrily", "sadly" or "gaily", and most of the information conveyed by the motion is contained in the adverb. The verb conveys the underlying desire, whereas the adverb conveys the intensity as well as the character's expectation as to how it will play out. If you see a character "walking" toward a water hole you infer that it is thirsty. If our character is walking "frantically" to the water hole, we infer that it is very thirsty. As it gets closer to the water hole and the frantic quality of the motion gives way to relief (note the intensity of the two emotions should be similar), we understand that the character believes that their great thirst will soon be satiated. This makes it all the more fun when the water hole disappears right before the character gets to it.

2.4 Putting It All Together

The challenge for the designer of a control system for an autonomous character is no different than that of an animator: an autonomous character's actions must be a direct and believable consequence of its character-specific desires and its character-specific beliefs about the effect of its action on the world so as to satisfy its desires.

However, as we have seen, observers expect that a character's beliefs about the world and the effect of its actions on the world will change as a result of its experience. For an animated character in a scripted animation, this is not a problem, it is sufficient that the character "act as if" it has revised its beliefs based on its observed previous experience with the world, and the animated character can rely on its animator to insure that that is the case. But for an autonomous animated character this is quite a problem because it must revise its beliefs in a way that makes sense to the observer. That is, it must learn and adapt its behavior (because it is only when its behavior changes that an observer can infer that it has learned something) in a manner consistent with the observer's expectations of what the character is able to learn. The role of the designer is to put the necessary scaffolding in place so the character can learn what it ought to learn.

Specifically, there are 2 fundamental kinds of things characters ought be able to learn.

1. They ought to be able to learn the immediate consequences of their actions, and the extent to which a given action is useful in satisfying a given desire. It may be useful because it satisfies the desire directly, or because performing the action brings the character closer to satisfying the desire. When performing the action it should be clear what they expect to happen.
2. They ought to be able to learn about their world, so that (a) they can choose the action that makes the most sense given the context in which they find themselves, and (b) they know whether objects in the world are good or bad and so can choose to approach or avoid them.

In the next section we will look at sources of insight and inspiration for understanding exactly how one would go about this.

3 Lessons

In this section we will look at 3 sources of inspiration: the current generation of digital pets, the domain of machine learning, and the domain of animal learning and training.

3.1 Lessons from Digital Pets

Simple learning has been integrated into several members of the current generation of digital pets, most notably AIBO and Dogz [12], Typically, learning is limited to biasing choice of action based on reward or punishment. Actions that appear to lead to reward increase in frequency whereas actions that appear to lead to punishment decrease in frequency. Despite the relatively limited amount of learning that the virtual pets can actually perform, the learning is believable and compelling. There are a number of reasons why the learning is so effective. In the case of Dogz, you can pat the dog with the mouse as a reward or squirt him with a spray bottle as punishment. Thus, the feedback signal is both simple and visible. The dog reacts to the feedback immediately and expressively suggesting that the consequences really matter. An immediate and observable change in the frequency of the behavior that precedes the feedback signal suggests that the dog associates the behavior with the good or bad consequences. This change in frequency together with the observed emotional response makes it appear that the dog learned from the experience. The entire model is very simple and intuitive. Finally, the creators of these digital pets can rely on our apparent innate tendency to read more into the behavior of autonomous creatures than may actually be warranted (thus is the power of the Intentional Stance.)

One moral from digital pets is that it doesn't have to be based on rocket science to work. Even the simplest and most limited form of learning, when done well, can be extremely compelling. Indeed, people tend to assume that digital pets learn a great deal more than they actually do.

3.2 Lessons from Machine Learning

Machine learning is an extremely active field with impressive results in a number of domains. As we will see however, traditional approaches need to be augmented if they are to be used for autonomous characters. For the purposes of this discussion we will limit our discussion to one type of machine learning called reinforcement learning. Excellent introductions to machine learning can be found in [1], [15].

Suppose a creature has a set of actions it can take, and a sensory mechanism that allows it to identify particular states of the world, and further there is a specific state of the world (i.e. a goal state) in which the creature receives reinforcement. Given that the creature finds itself in a state other than its goal state, the problem facing the creature is to find the "best" sequence of actions that will get it to its goal state. The problem comes when the creature doesn't know the consequences of its actions, i.e. what state will result if it takes a given action in a given state. In this case, there is nothing for it but to learn the relationship between states, actions and consequences. This is the problem that reinforcement learning addresses.

State refers to a specific, and hopefully unique, configuration of the world as sensed by the creature's sensory system. As such, *state* can be thought of as a label that is assigned to a sensed configuration. The space of all possible sensed configurations of the world is known as the *state-space*.

Action is what the creature does so as to interact with the world. Performing an action is how a creature changes the state in which it finds itself. Typically when reinforcement learning is discussed in relation to autonomous creatures, a creature is assumed to have a finite set of actions of which it can perform exactly one at any given instant. The set of all possible *actions* is referred to as the *action-space*.

The creature receives reinforcement when it reaches a state in which it can satisfy a specific goal. For example, if a dog sits and gets a treat for doing so, the reward or reinforcement is the resulting decrease in hunger or pleasure in eating the treat.

Reinforcement learning sets out to learn the best sequence of actions for the creature to take in order to get it to its goal state given that it finds itself in arbitrary *state s*? That is, if the creature is capable of performing n actions and it is in *state s*, which one of those *n actions* is the best action to perform so as to move it closer to its goal state? A common way to conceptualize the problem is to think of a big table, in which each row represents being in a given *state*, and each column represents a given *action*. Each entry in the table is called a *state-action pair*, and the goal of the learning algorithm is to learn a value for each state-action pair that reflects its "utility" with respect to the goal. Once all of the values for the state-action pairs are learned, it is a simple matter to make use of the table. If in *state s*, find the *state-action pair* for *s* that has the highest value and perform that *action*. The trick, of course, is to learn the appropriate value for each *state-action pair*.

Under certain conditions it turns out to be remarkably straightforward to learn the appropriate values for the *state-action pairs*. Christopher Watkins developed an iterative technique called Q-learning that does just that. Watkins called the learned value of a state-action pair its *Q-value*. What is startling about Q-Learning is that Watkins was able to show that by using only a local update rule the system could learn the optimal value for each of its Q-values [15]. The process of updating the value is known as *credit-assignment*.

Thus, in Q-learning the creature is placed in an arbitrary state and it moves through its state space by performing actions until it arrives at its goal state. It chooses what action to perform next by mostly using the policy of choosing the action that has the highest Q value for its current state. As it moves from state to state it performs credit assignment as described above. When the creature reaches its goal state, the process is repeated starting in another state. Watkins showed that if this process was repeated a sufficient number of times for all state-action pairs, the Q-value for each state-action pair would approach a value reflecting its optimal discounted utility with respect to the goal state [15]. Thus learning is fundamentally a process of exploration.

While reinforcement learning provides a theoretically sound basis for building systems that learn, there are a number of issues that make it problematical in the context of autonomous animated creatures. None of these issues is insurmountable but they do need to be considered. The more important of these issues include:

1. **Representation of state**: For any but the most toy problem, the state-space can quickly become huge, even though most of it is irrelevant. Consider a dog that is to be taught to respond to arbitrary acoustic patterns. The space of all possible acoustic patterns is (a) continuous and (b) far too big to permit an exhaustive search even if it was discretized. Of course, the fact that most acoustic patterns are irrelevant to most dogs suggests that it isn't necessary to represent all possible acoustic patterns a priori but rather it is sufficient to discover, based on experience, those acoustic patterns that seem to matter and add them dynamically to the *state-space*. This process is known as *state-space discovery* and is an essential component to successful learning in the real world.
2. **Representation of action**: Q-learning assumes that the form of the action remains constant. However, for an animal or animated character, the form of the action matters almost as much as the choice of action. To get around this problem, one could "discretize" the action-space, e.g. have 10 different actions each of which correspond to a specific style of walking. This, however, would cause the search space to grow substantially. Indeed, for a creature that can recognize 100 individual states of the world, each action adds 100 state-action pairs that must be visited repeatedly in order to learn their optimal Q-value. In addition, if the creature needs to learn actions that weren't programmed in (e.g. to learn that a novel trajectory such as figure-eight is an action that lead to reinforcement), it needs to perform the equivalent of state-space discovery in action space, i.e. *action-space discovery*.
3. **Representation of time**: Q-learning makes the assumption that all actions take 1 unit of time to complete and that credit assignment occurs every time step. In fact, actions in an animal or an autonomous character take variable amounts of time to complete. A "sit" may take a second, whereas a "fetch ball" may take 10-15 seconds. Even the same action may take a variable amount of time. For example,

the time taken to complete "fetch ball" depends a great deal on the distance one throws the ball.

4. **Multiple goals**: Creatures & characters have multiple goals/desires that they attempt to satisfy. Most work on learning assumes a single goal. For example, to use Q-learning for a character with multiple goals, one would need a separate Q-table (table of state-action pairs) for each goal and a way of choosing between which goal to attend at any given time.
5. **Learning vs. behaving**: Learning is just one thing a character needs to do. For animals and characters alike, learning augments existing behavior. Most approaches to machine learning assume that (a) learning is the primary task of the system and (b) learning starts with *tabla rasa*. Conversely, relatively little work has been done in developing behavior architectures in which learning can take place as part of the overall agenda of the character or creature.
6. **Exploration**: As the size of the state and action spaces grows it becomes critical to have strategies or heuristics in place to guide the creature's exploration of its state-action space. That is, to experiment with those state-action pairs that are most likely to be ultimately valuable. In most approaches to machine learning this is left as "an exercise for the reader" since researchers are more concerned with asymptotic performance rather than initial performance. By contrast, animals and characters are probably more concerned with quickly learning "acceptable" solutions than "optimal" solutions. [7], [14], [9]

Conceptually, animals face the same problems as those faced by machine learning systems, but appear to have no problem learning what they ought to learn. The question becomes how they do this. While the real answer is we don't know for sure, careful study suggests that the answer may lie in the use of heuristics and built-in structure that have the effect of simplifying the learning task. Indeed, the process of training is really one of guiding the animal's exploration of its state and action space toward the performance of specific actions in specific contexts.

3.3 Lessons from Nature

In this section we will review some of the key lessons to be gained from animal learning and training. See [7], [9], [10], [14], [16] for wonderful discussions of animal learning and training.

3.4 Learning

In nature, learning is a mechanism for adapting to significant spatial and temporal aspects of an animal's environment that vary predictably, but at a rate faster than that to which evolution can adjust, or which can not be coded for in the genes. Indeed, the most adaptive course in this case is to evolve mechanisms that facilitate learning these rapidly varying features [11]. Thus, evolution determines much of what can be learned and the manner in which it is learned. Often this takes the form of innate structures that have the effect of dramatically simplifying the task of learning specific things [7], [11]. Some of these important heuristics include:

- **Role of Variability**. Variability of action and context is absolutely essential to learning. Variability of action allows the animal to discover new causal associations, and by varying how an action is performed find the most reliable form of the action. Variability of context allows the animal to identify relevant cues to apparent causality, and in particular identify those cues that increase the reliability of the association between an action and an outcome. Animals appear to be sensitive to the variability of the expected outcome [14]. Indeed, the more variable the outcome, the more variable the choice & form or action. Trainers often make use of these phenomena by varying the reward associated with the performance of a desired trick. The computational implication is that the choice and style of action should be probabilistic but biased toward style toward those choices and styles of actions that lead to good consequences or avoid bad consequences.

- **Role of Motivation.** In general, the more motivationally significant the consequences the more rapidly the animal will learn the context and actions that seem to lead to it. In some cases, the motivation is an end-result such as a treat, in other cases, it is the chance to perform an activity, and in still other cases, the action itself is rewarding [9]. [11], [14]. The point, however, is that learning and motivation are closely linked.
- **Frequency of Action is Proportional to its Perceived Consequences.** Actions that seem to lead to good things tend to be expressed more often than those that don't (this is known as Thorndike's Law of Effect) [14], [7]. This behavior makes sense for two reasons. First, it increases the chances of a desired outcome. Second, by increasing the frequency of a "promising" action, but varying how it is performed, they are in effect exploring a potentially promising neighborhood. This has three important implications for our computational model. Firstly, there needs to be some representation of consequences, both good and bad. Second, as mentioned earlier the probability of a given action should reflect the value of the likely consequences of performing that action given the state of the world at that moment. Third, the focus of learning should first be on learning the likely consequences of actions, and then learning the contexts in which the action is especially reliable at producing the desired consequence.
- **Animals constrain their search for apparent proximate causality to a small temporal window around the performance of an action**. The rule of thumb in dog training is that unless the consequences of an action are signaled within 2 seconds of the performance of the action, a dog won't learn the connection [10], [14]. Similarly, events that occur within a small temporal window preceding and perhaps overlapping with the performance of the action are assumed to represent the candidate set of stimuli relevant for increasing the reliability of the action. The computational implication is that our system needs to maintain memory sufficient to be able to answer questions such as "what stimuli were active within a given temporal window of an action becoming active." The good news is that the temporal window can be relatively short. Similarly, the relevant consequences are those that immediately follow the completion of the action.
- **Time and Rate are Fundamental Building Blocks**. Animals act as if they have internal representations of time, quantity and rate and are capable of using these representations to make commonsensical decisions about how to organize their behavior [6] Time, rate and quantity should be explicitly represented in the system and used not only to guide choice of action, but exploration as well.

As we will see in the next section, the difference between a great trainer and a mediocre one is the degree to which the trainer takes advantage of the heuristics that seem to guide learning in animals.

3.5 Training

It is useful to examine training techniques for two reasons. First, to understand what the techniques imply about how animals learn. Second, the techniques themselves may be useful as a way to train autonomous characters. At a very basic level the trainer must help the animal answer five questions:

- **Why do it?** The trainer must insure that the consequences are motivationally significant to the animal; otherwise, it is unlikely the animal will be motivated to learn. That is, it must be clearly significant relative to the inferred desires of the animal.
- **What to do?** The trainer must signal the animal when it has performed the desired action that is causal to the subsequent reward appearing. Reflecting the narrow window that animals appear to use for inferring causality, typically trainers use an event marker such as a click or a whistle. The click, that has been previously associated with a reward, acts both as an event marker as well as a bridge between the end of the desired behavior and the delivery of the reward [11].
- **How to do it?** Good trainers are as sensitive to the form of the motion as a good animator. Varying the level of reinforcement is one way a trainer can cause the animal to vary its performance of the behaviour, since animals seem sensitive to variations in outcomes [Gary Wilkes, personal communication], [11]. By rewarding ever-closer approximations to the desired final form of the behaviour, a process known as *shaping*, the trainer effectively guides the exploration. In the case of behaviours that occur infrequently, the trainer may *lure* an animal into performing an approximation of it — for example, by moving a piece of food over the dog's head, a dog may be lured into sitting down. Indeed, virtually all animal tricks start with a naturally occurring action that is then shaped and perhaps expressed in a novel context [9].
- **When to do it?** Typically, trainers will only begin associating a cue or context for a behaviour once they are sure that the animal has learned the desired behaviour [Wilkes, personal communication][11]. They associate the cue by issuing it as the animal is beginning to perform the desired behaviour. By rewarding productions of the behaviour when this cue is given, and ignoring instances when it is not – the animal responds by decreasing the spontaneous production of the behaviour and begins to respond to the cue by performing the behaviour. This process typically takes between 20 and 50 repetitions [Wilkes, personal communication].
- **How long to do it?** Trainers rely on the seeming ability of animals to learn intervals, in this case the expected interval and variance between the onset of the behavior and the subsequent reward.

The key point here is that the role of the trainer is to simplify the world for the animal by guiding the animal's exploration of its state and action space. By doing so they effectively address one of the major problems faced by machine learning systems, namely how to search the state & action spaces intelligently. As we saw in the section on animal learning, animals also use heuristics such as tight temporal windows for inferring causality and variance of reward to guide their own exploration. These ideas have influenced our approach to learning, as we will see in the next section.

4 Our Approach

The learning mechanism in C4, the toolkit developed by the Synthetic Characters Group for modeling autonomous animated characters, incorporates many of the lessons discussed above. Space does not permit a detailed discussion of the specific learning mechanism (see [3] and [7] for a detailed description of C4 and the actual learning mechanism) but the key lessons we have incorporated into the architecture are summarized below.

- We exploit aspects of the world that effectively limit the search space. For example, temporal proximity is used to infer apparent causality. That is, we utilize a temporal attention window that overlaps the beginning of an action to identify potentially relevant state. Similarly, we assign credit to the action that immediately precedes a motivationally significant event in a manner similar to Q-Learning [1].
- We utilize loosely hierarchical representations of state, action and state-action space, and use simple statistics to identify potentially promising areas of the respective spaces for exploration. Through a process known as innovation we grow the hierarchy downward toward ever-more fine-grained representations of state and more specific, and hopefully more reliable state-action pairs. This approach was inspired in part by [5]. Thus, the process is one of starting with rather generic state-action pairs, and generating more specific instances of pairs for which there is some evidence that they are both potentially valuable and could be made more reliable by being more specific as to the context (i.e. the state) in which they are performed. Measures of novelty and reliability are used to guide this process as well as temporal proximity.
- We use natural feedback signals to guide exploration. For example, a significant change in a motivational variable is a natural reward signal for determining the value of state-action pairs (or their equivalent), but also for guiding state-space and action-space discovery. For example, if a model-based recognizer is being built from examples to identify a particular state of the world (e.g. an acoustic pattern that signals when the dog should beg), we use the reward signal to disambiguate between good and bad examples. A good example is one that leads to an expected reward. As the state and action-space trees grow downward, we are then able to make use of the new states and actions in our state-action tree as described above.
- We utilize biases that affect the frequency and timing of actions. These biases take two forms. One bias is to perform actions that have led to reinforcement in the past. This not only allows the creature to exploit what it knows but also more opportunities to discover more reliable variations. There is also an innate bias to perform a given pattern at a given time, thereby providing an opportunity to incorporate the pattern into the behavioral repertoire should it prove useful.

- We tie variability of action to variability of outcome. That is, the variability in expected outcome is treated as a signal indicating the degree to which the creature is successful in controlling its environment through its actions. When the outcome is highly variable the choice of action, and form of action is highly variable as well.

The learning mechanism in C4 is still a work in progress but is already at a level at which we can train a virtual dog using techniques borrowed from real dog training. In the following section we discuss some of the characters that have been built incorporating this approach to learning.

5 Our Experience So Far

In this section we review a number of the characters that we have built to date using this approach.

5.1 (Void *): A Cast of Characters

In (Void*): A Cast of Characters, there are three characters hanging out in a late night diner. A human participant can "possess" a character by manipulating an instrumented set of buns and forks, and force the possessed character to perform a variety of dance steps. Each character responds to possession and dancing differently, based on their innate personality, past experience and motivation. Depending on how the user controls the interface, the characters can have a fun time dancing or they can have painful experiences falling down on the floor, and they update their belief about the desirability of being possessed accordingly. In either case, the attitude toward being possessed is reflected in the quality of their dancing, as well as their facial expression. Once a character is un-possessed, it may continue dancing or go back to its seat by its own free choice. The strength of the desire to dance is determined by the overall affective feedback from the previous dancing experience. If it decides to continue to dance, it dances using those steps that seemed most popular with the participant (i.e. those that had the highest frequency of being chosen.)

While the learning in (void *) was very simple, participants found it both compelling and believable. It was easy for them to see the observable change in behavior that resulted from the character's learning whether being possessed was a good or bad thing.

5.2 Duncan

Duncan, the Highland Terrier is an ongoing research effort to build an autonomous animated dog whose ability to learn, behavioral complexity and apparent sense of empathy rivals that of a real dog. Duncan has been featured in two significant projects to date. One project is *sheep|dog* an interactive installation piece in which a user plays the role of a shepherd who must interact through a series of vocal commands with Duncan to herd a flock of sheep. This system demonstrated some of the basic reactive, perceptual and spatial abilities of Duncan, as well as his ability to classify user utterances as one of six possible commands. This classification could be trained through a “one-shot learning” interface so that a new user could achieve a high recognition rate after a very short (about 15 seconds) training routine.

The learning algorithms being developed by our group were put to use in the second project, named *Clicker*, in which a user can train Duncan using “clicker training” as described earlier. In this simulation, Duncan can be trained to associate vocal commands with behaviors, and demonstrates a number of the phenomena that one sees in real dog training (i.e. Thorndike’s Law of Effect, shaping, resistance to extinction etc.) Given an initial repertoire of a dozen basic behaviors (e.g. “sit”, “shake”, “lie-down”, “beg”, “jump”, “go-out”) together with basic navigational and behavioral competencies we have been able to train him to respond to both symbolic gestures (i.e. game-pad button presses), and more significantly to arbitrary acoustic patterns (indeed one user trained Duncan to respond to commands in Gha, the language of Ghana). A dozen such tricks can be trained in real-time within the space of 15 minutes. We have also demonstrated simple shaping with both motor systems and complex shaping and luring. Duncan is also capable of simple spatial learning, and has an understanding of object-permanence.

5.3 Goatzilla

Goatzilla is a dinosaur with the “brain of a dog”. While Duncan’s model of learning was very much inspired by the requirements of “clicker training”, the model of learning incorporated in Goatzilla was inspired by the work of Gallistel who has proposed a model of learning in animals that seems to explain a wide range of learning phenomena from classical to operant conditioning. Rather than relying on an associative explanation, Gallistel proposes that the animal learns the temporal interval between events, and the rates of occurrence, and uses this information in conjunction with some simple heuristics to infer causality in the world. In our case, Goatzilla learns to satisfy his drives by using a form of time-rate learning to discover important and relevant causality in his world. Note: this model encompasses what is needed to support clicker training, and so we have updated Duncan to utilize this approach as well.

Goatzilla seeks to understand the cause of motivationally salient stimuli and learns to expect their onset. When he is surprised by the appearance of a stimulus, he forms and subsequently tests a hypothesis that is meant to explain the event. The hypotheses created might involve self-action (what I observed was the result of something I did), or, it might involve other salient stimuli observed by the creature (stimulus A tends to precede stimulus B by time t). After being tested, these hypotheses give rise to future expectations and potentially expectation violations, which will be used to help improve each hypothesis by refining its context. In the absence of salient stimuli, Goatzilla is motivated by a curiosity drive to experiment, both by testing out uncertain hypotheses he has formulated, and also by experimenting with objects and actions he is unfamiliar with.

The result of this is that Goatzilla can learn that (a) eating is a good thing to do when hungry, (b) sheep are a good thing to eat, (c) if no sheep are present, kicking the shed in which the sheep live is a good strategy for making sheep appear.

Because the hypotheses allow the creature to predict the consequences of actions, the predictions work in conjunction with its drives to adjust the creature's affect when these events occur. Thus it provides the necessary framework for the beginnings of an emotion system.

5.4 α_{wolf}

α_{wolf} is an installation we will be showing at Siggraph 2001, that allows multiple participants to explore the social dynamics in a simulated wolf pack. By “growling”, “whining”, “barking” or “howling” into a microphone, and by directing their wolf pup to interact with another pack member, a participant helps their wolf pup find its place in the social order of the pack. Based on its interactions with the other wolves, the wolf pup learns whether it is dominant or subordinate with respect to each of the other wolves in the pack and adjusts its behavior accordingly. The wolf’s emotional state is influenced by its level of certainty with respect to its interactions with others. While a wolf may prefer to be the alpha wolf, what it really cares about is the extent to which it has reliable strategies for avoiding dangerous conflict.

6 Conclusion

The computational model underlying all our work focuses on the kind of learning that dogs do, and it must be said that not only do dogs learn many more things than are addressed in this model at present, but also that people can learn a great deal more than dogs. Nonetheless, we believe that this kind of everyday learning underlies much of our everyday common sense, and our expectation about what a sentient creature is minimally capable of learning. As people will interact with synthetic characters over extended periods of time they will expect them to learn from experience in the same commonsensical way. In the long run, only Wile E. Coyote can get away with not learning from experience.

Acknowledgements. The work reported in this paper represents the collective work of the Synthetic Characters Group at the Media Lab of MIT. Members of the group whose ideas are reflected here include: Matt Berlin, Robert Burke, Marc Downie, Scott Eaton, Jesse Gray, Michal Hlavac, Damian Isla, Yuri Ivanov, Michael Patrick Johnson, Christopher Kline, Delphine Nain, Bill Tomlinson, Ben Resner, Ken Russell, and Song-Yee Yoon. Thanks also to Gary Wilkes for his insights into dog training and behavior. This work is sponsored by the Digital Life Consortium of the Media Lab. Finally; the author would like to express his thanks to the organizers of the International Conference on Virtual Storytelling for the opportunity to present this paper.

References

1. Ballard, D., An Introduction to Natural Computation, MIT Press, Cambridge 1997.

2. Blumberg, B.M., Gaylean, T., Multi-level Direction of Autonomous Creatures for Real-Time Virtual Environments, *SIGGRAPH 95 Conference Proceedings*, 1995
3. Burke, R., Isla, D., Downie, M., Ivanov, Y. and Blumberg, B., Creature Smarts: The Art and Architecture of a Virtual Brain. *Proceedings of the Computer Game Developers Conference*, San Jose Ca., 2001
4. Dennett, D. The Intentional Stance. Cambridge, Mass.: MIT Press. 1987.
5. Drescher, G.L. Made-Up Minds:A Constructivist Approach to Artificial Intelligence, MIT Press, Cambridge 1991
6. Gallistel, C. R., & Gibbon, J. Time, rate and conditioning. Psychological Review' 107, pp289-344.
7. Gould, J & Gould C. The Animal Mind. W.H. Freeman, New York, N.Y. 1999.
8. Isla, D., Burke, R., Downie M., and Blumberg B. A Layered Brain Architecture for Synthetic Creatures. *Proceedings of the International Joint Conference on Artificial Intelligence*, Seattle Wa., 2001.
9. Lorenz, K. The Foundations of Ethology. Springer-Verlag, New York, NY. 1981.
10. Plotkin, H., Darwin Machines and the Nature of Knowledge. Harvard University Press, Cambridge Ma. 1994.
11. Pryor, K. Clicker Training for Dogs. Waltham, MA.: Sunshine Books, Inc. 1999.
12. Resner, B., Stern A., Frank, A. The Truth About Catz and Dogz. *Proceedings of the Computer Games Developer Conference*, San Jose, Ca., 1997.
13. Rose, C.F., Cohen, M., Bodenheimer, B., Verbs and Adverbs: Multidimensional Motion Interpolation - *IEEE Computer Graphics And Applications*, Volume 18, Number 5, 1999
14. Shettleworth, Sara J. Cognition, Evolution and Behavior. New York: Oxford University Press. 1998.
15. Sutton, R. and Barto, A. Reinforcement Learning: An Introduction. MIT Press: 1998.
16. Thomas, F. and O. Johnson. 1981. *The Illusion of Life: Disney Animation.* New York: Hyperion.
17. Yoon, S., Blumberg, B.M., Schneider, G.E., Motivation-Driven Learning for Interactive Synthetic Characters, *Autonomous Agents 2000 Conference Proceedings*, 2000

Real Characters in Virtual Stories

Promoting Interactive Story-Creation Activities

Isabel Machado[1], Ana Paiva[2], and Paul Brna[3]

[1] ISCTE, INESC and CBLU — University of Leeds,
Rua Alves Redol 9, 1000 Lisboa — Portugal
Isabel.Machado@inesc.pt

[2] IST-Technical University of Lisbon and INESC,
Rua Alves Redol 9, 1000 Lisboa — Portugal
Ana.Paiva@inesc.pt

[3] Computer Based Learning Unit — University of Leeds,
E.C. Stoner Building, LS2 9JT Leeds — England
P.Brna@cbl.leeds.ac.uk

Abstract. Interactive experience in a virtual world. We take the line that children need to be both engaged in the action through role play and given the opportunity to reflect on the significance of their actions to understand something of their significance in terms of both the narrative and its ethical significance. This requires a system that incorporates the children's actions into the unfolding plot. We introduce the Support And Guidance Architecture (SAGA), a plug-in architecture for guiding and supporting children's interactive story creation activities. This is illustrated with reference to Teatrix, a collaborative virtual environment for story creation, which provides the children with the means for collaboratively creating their story on a virtual stage.

1 Introduction

In the past years, several researchers from distinct and different areas such as interactive drama [4], computer-games (*e.g.*: SIMS, Shadow of Memories, etc.) have tried to develop a system that would provide its users with an interactive experience, and simultaneously would allow the users to act out a role within that experience. However, none of these researchers point to a clear solution that establishes a compromise between plot, characters and users. By placing the user (or player) inside the plot, it is necessary to accommodate the actions he/she takes into the unfolding plot, and, at the same time, to guarantee the achievement of a true interactive experience. The role of the user has changed from being just a spectator of the story to a first person character — for example, in the *Shadow of Memories* game, the player can be the detective of his own murder, discover the assassin and avoid being killed. Nevertheless, stories are not only presented to users in the format of interactive computer systems, they add color to our lives since our early childhood. Cognitive development theorists and psychologists ([10], [14]) suggested that through make-believe activities children

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 127–134, 2001.

start to understand the mysterious world they live in, and that in a fantasy scenario they engage in new experiences. By doing this, they acquire proficiency in acting in the real world. In middle-childhood, fantasy takes the form of board, video and computer games, and in creative drama and theatrical performances on the school premises as well. Also, at this developmental stage, children prefer rule-based games, in which they tend to create their own rules or even use them to provide an arena in which to compete [12].

These two streams of research have led us to the problem of developing a system that would be able to convey a fulfilling interactive story to its users (children) and, at the same time, to allow such users to play inside the story as characters. Our approach to this problem, it is the development of a general architecture — *Support And Guidance Architecture (SAGA)* that can be used in different collaborative story creation applications. The aim of *SAGA* is to provide such applications with a mechanism to give support and guidance to children during the story creation process. Additionally, the research goal is to provide the children with an interactive story, where they have a character to control in the story, and at the same time they have the opportunity to reflect upon their characters' actions. With this reflection activity we aim at providing the children with a psychological portrait of the story characters, which may contribute to a "better" story achievement.

2 *SAGA*: Support and Guidance Architecture

The development of *SAGA* was based on the assumption that the story creation process is composed of two distinct phases: story definition/preparation and story construction (*de facto*). In the first phase the children define the basic elements for their story: the cast and setting, and in the second they collaborate between themselves to build their story. Therefore, both phases are dependent on the story creation application that is using the services of *SAGA* (for example, we could have applications that provide the children with the means to create their story in the format of a play, a cartoon, etc.).

2.1 Concepts

As other researchers have done before us ([13]), we decided to adopt Vladimir Propp's morphology [11] as the underlying theory of narrative for our model. However, due to the lack of interactivity associated with this theory, we decided to enrich it with some AI concepts and also some educational practices. The major concepts, derived from the work of Propp, present in the architecture are:

- story — is a sequence constituted by the 31 functions. A story must be started by a villainy or a lack, and proceeds through intermediary functions to a reward, a gain or in general a liquidation of misfortune;
- function — which can be understood as both the actions of the characters and the consequences of these actions for the story;

– role — a set of behaviours (specified by a set of functions) that are known to both the characters and audience [3].

Additionally, we defined the concept of an actor, which emerged from the study of theatrical performances. An actor is the physical representation or appearance of a character (example: a young girl, a wolf, a witch, etc.). And finally, a character is the conjunction of two different concepts: an actor and a role. The character is the one that acts in the story, accordingly to its role and in the skin of the actor[1]

2.2 Integration

To integrate *SAGA* into a story creation application, it is necessary that the latter complies with two important properties: (1) to be observable — *SAGA* must be able to inspect the state of the application and also, (2) to allow changes — *SAGA* must be able to introduce new elements into the story creation application or even to take some actions in the story creation application.

2.3 Components

The components of *SAGA* are: the *Facilitator*, the *Scriptwriter*, the *Director Agent*, the *Narrative Guidance Engine* and the *Reflection Engine* (see Figure 1).

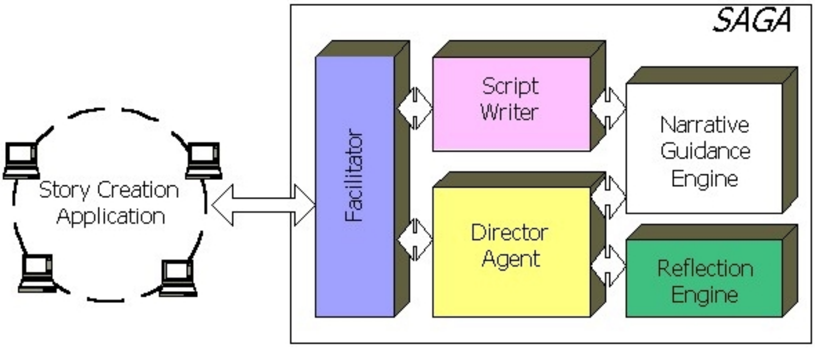


Fig. 1. *SAGA*

The *Facilitator* is the component of *SAGA*, which establishes the bridge between the architecture and the story creation application.

[1] In fact, Propp indicates that an actant can have more than one associated function. Actants however are not exactly the same as actors.

The *Scriptwriter* has the main goal of building an initial story situation in accordance with the story elements previously chosen by the children. The definition of the story's initial situation was based on the work of Propp, and can be specified as the situation in which the characters are introduced, the relations between such characters are established, and the story is situated in terms of temporal and spatial location. To do this, a set of templates is available each of which is defined by a set of minimum requisites that must be satisfied to have an initial story situation. To better illustrate these ideas, a template fragment is presented:

```
<violationFunction> <text> However, now that she is alone in the
world and becoming older, she is thinking more and more to
explore the </text><sceneName>forest</sceneName> <text>and find
the enchanted lake. If she could find such lake she would have
money to try finding her parents.</text> </violationFunction>

<firstVillainAppearanceFunction> <text> Also, in the </text>
<sceneName> forest </sceneName><text>lived a terrible and
feared</text> <villainType>wolf </villainType><text>named,
</text><villainName>Herman</villainName><text>, who was the
guardian of the magic lake. </text>
</firstVillainAppearanceFunction>

<reconnaissanceFunction> <text> In her deeper thoughts
</text><heroName> Mary </heroName><text> wanted to know more about
this lake guardian. How was it? Was it so bad as his mother
said?</text> </reconnaissanceFunction>
```

In the above template, we can distinguish the parameters (<villainName>, <heroName>, etc.), that can be instantiated with distinct values, which provides the possibility of generating different initial story situations. From the template excerpt we can see that any stories that are created using it will evolve around the search for the *guardian of the magic lake* and its treasures, but everything else is left open to children's creativity.

Although, the motive can be established at a large-grain size, by the type of story template, there is also the need to establish a set of challenges to be discovered throughout the story progression. These challenges are intended to enhance the story with an extra degree of suspense, which would be translated into a more interesting experience for the children, in a game like way [15].

The main goal of the *Narrative Guidance Engine* is to generate the space of all plot points for a particular story. A plot point is an important story situation, which should be played by the children in order to achieve the goal of the story (similar to the approach taken in the OZ project [7]). These plot points are defined from the initial story situation and from the functional roles performed by the characters. The space of all plot points is the result of all paths between plot points that make possible the achievement of the end of the story, implicitly reaching also the goal of the story. An evaluation function was therefore defined to determined, at each point in time, which is the best path to follow.

The *Director Agent* is the component that has the responsibility for deciding how and when some particular kind of support should be provided. To do this, we are developing the concept of a narrative agent which is equipped with a decision process that is used to consider what to do. This decision process is performed with the help of the agent's narrative memory. Its narrative memory is organised in the form of episodes and contains information about story progression from each story character's point of view. Each episode is constituted by three important events: crisis, climax and resolution [2]. In the end of the story creation activity the narrative agent can use the various character-centred stories, stored in its memory, to generate a unique story that reflects the overall experience of the story creation activity. The *Director Agent* also has the important role of asking a child to reflect upon the actions performed by her character. This can happen, for example, when it detects a conflict between the actions performed by the child and her character's role.

The *Reflection Engine* is the component that on demand (by the Director Agent) generates a reflection moment. The idea is that a child is asked to put herself in someone else's shoes and explain the meaning of her character's current behaviour [1]. Additionally, all the other children collaborating in the story creation process should be informed about such reflection, since it can influence the flow and development of the story.

With this component we aim at providing the children with the opportunity to inspect the characters' minds, and understand their behaviours and motivations. By doing this they have the opportunity to act (by means of their characters) in accordance with such behaviours and even explore more deeply the plot of the story (for example: if a child see the character wolf running after her girl character, and she knows that the wolf is hungry so maybe it is only starving and not bad).

3 Application of *SAGA*

Teatrix is an application developed under the NIMIS (Networked Interactive Media In Schools) project, which was an EU-funded project under the Experimental School Environments (ESE) program. It is a collaborative virtual environment for story creation, which provides the children with the means for collaboratively creating their story on a virtual stage. The children are able to create the stories using a set of pre-defined scenes and characters. These characters may act on behalf of the children or autonomously (for further details see [6]). To act and create the story the children have a set of actions and props available (see Figure 2).

The application of *SAGA* in Teatrix, starts by providing the children with an interesting initial story situation. The children are given an introduction to the story, similar to what happens in a game, but unlike the games everything else about the plot is left to be determined by the children. The role of *SAGA* is to

Fig. 2. Teatrix

guide and support the progression of the story. To do this, *SAGA* has the power to introduce new props or characters into the virtual world, and also to interact with the children through the reflection moments (implemented in Teatrix, in a reflection tool called *Hot-seating* [5]).

At each point of the story, *SAGA* provides the possibility of confronting the children with what is being done and what will be needed in order to accomplish the goal of the story (mapped in the space of story plot points). For example, take into consideration a story being created from the template presented above:

– *situation*: Mary meets Herman at the forest
– *next plot point*: struggle between hero and villain
– *actions:*
  • Mary: talk with Herman and offer him a mushroom;
  • Herman: accept the offer;
– *Director's perspective*: a conflict occurs between the role and the behaviour of the villain character, which means that the a reflection moment is must be triggered;
– *reflection moment*: the *Hot-seating* interface appears in the monitor of the child controlling the wolf and she has to justify why her character is not performing accordingly to its role. At this point, if the child decides to change her character's behaviour the plot point may be achieved, or in the opposite

situation the *Director Agent* may take a different course of action and introduce a new villain in the story, assuming that Herman is now assuming the role of a helper.

This is just an example, of how the architecture is integrated within *Teatrix*, and the story would evolve until the final goal has been reached.

4 Preliminary Results

After a few tests of the architecture itself, we came to the conclusion that the space of all plot points generated were directly proportional to the complexity of the initial story situation, and that the majority of the plot points are useless for the story. From this empirical result, we decided to start doing the generation in a phased way, i.e., by dividing the basic structure of our model into phases and generating the paths not only according to the initial story situation but also by considering the part of the story already achieved.

Also from a preliminary integration results with *Teatrix*, we concluded that the introduction of the reflection engine, in the form of *Hot-seating*, has been accepted well by the children, who demanded a higher degree of control over their characters. Also, we have evidence that some of the reflection moments have been referenced inside the story.

5 Conclusions and Future Steps

In this paper we have proposed a plug-in architecture for guiding and supporting children's interactive story creation activities. However, when doing such guiding, *SAGA* also has to ensure the 3 requisites for providing each child with an engaging interactive experience ([9], [8]): (1) immersion — by making it possible for the child to feel herself part of the story and with the power to act in it; (2) agency — by taking into account her actions as a contribution to the flow of the story; (3) and, transformation — by making it possible for the child to put herself in someone else's shoes and in this way to explore a multitude of different situations.

On the one hand, we argue that the usage of *SAGA* enhances the story environments with a support and guidance strategy, but on the other hand it will support the user's need for an interactive and engaging experience.

Acknowledgements. We thank the other members of the INESC team involved in the design, development and evaluation of Teatrix application.

References

1. G. Bolton. *Acting in classroom drama : a critical analysis.* Stoke-on-Trent : Trentham, 1998.

2. L. Egri. *The art of dramtic writing - Its basis in the creative interpretation of human motives.* Touchstone Book - Simon and Schuster, 1960.
3. B. Hayes-Roth. Acting in character. In R. Trappl and P. Petta, editors, *Creating Personalities for Synthetic Actors.* Springer Verlag, 1997.
4. B. Laurel. *Computers as Theatre.* Addison-Wesley, 1993.
5. I. Machado and A. Paiva. The child behind the character. In *Working Notes of the Fall Symposium Series - Socially Intelligent Agents: The Human in The Loop.* AAAI Press, 2000.
6. I. Machado, R. Prada, and A. Paiva. Bringing drama to a virtual stage. In *Proceedings of Collaborative Virtual Environments Conference.* ACM Press, 2000.
7. M. Mateas. An oz-centric review of interactive drama and believable agents. Technical Report CMU-CS-97-156, Carnegie Mellon University, 1997.
8. M. Mateas. A neo-aristotelian theory of interactive drama. In *Working Notes of the Spring Symposium of AAAI 2000 on Artificial Intelligence and Interactive Entertainment.* AAAI Press, 2000.
9. J. H. Murray. *Hamlet on the Holodeck - The future of the narrative cyberspace.* The Mit Press, 1998.
10. J. Piaget. *Play, dreams and imitation in childhood.* London: Heinemann, 1951.
11. V. Propp. *Morphology of the folktale.* Austin: University of Texas Press, 1968.
12. D. Singer and J. Singer. *The House of Make-Believe.* Harvard University Press, 1990.
13. N. Szilas. Interactive drama on computer: Beyond linear narrative. In *Working Notes of the Fall Symposium Series - Narrative Intelligence.* AAAI Press, 1999.
14. D. Winnicott. *Playing and Reality.* Basic Books, 1971.
15. M. Young. Creating interactive narrative structures: The potential for ai approaches. In *Working Notes of the Spring Symposium of AAAI 2000 on Artificial Intelligence and Interactive Entertainment.* AAAI Press, 2000.

Real-Time Character Animation Using Multi-layered Scripts and Spacetime Optimization

Eric Menou, Vincent Bonnafous, Jean-Pierre Jessel, and René Caubet

IRIT-UPS 118 route de Narbonne, 31062 Toulouse Cedex 4 France
{menou|bonnafou|jessel|caubet}@irit.fr

Abstract. We introduce a new animation system dedicated to real-time character animation. A multi-layered script system is used to control the motion from the higher semantic level (used to produce scenarios using high level orders) to the geometrical aspect (required to control the movement in a precise way). Two low-level animation subsystems are proposed, depending on the requirements of the final user. One uses a blending layer to mix the tasks generated by the script system, whereas the other performs on-the-fly spacetime optimization to compute the resulting motion. Two applications are described, each one using one of these two systems, putting in evidence their respective advantages and drawbacks.

1 Introduction

Characters are at the center of number of virtual storytelling applications. Few stories indeed don't imply at least one character, whether it is humanoid or not. Animating characters remains however a recurrent problem as much as for real-time applications (such as virtual environments or video games) as for offline production. The difficulty often comes from the fact that finding a good balance between automation and control is not a trivial task [1]. Automation, that generally uses procedural methods (dynamics simulation for example), allows to generate more realistic animations (in the physical sense) and thus more credible, but increasing the automation means decreasing the control of the animator, and tends to produce uniform motions. If the entire animation is made by hand by an animator, he can express more creativity and originality and make show through the feelings of persons. But not only this implies a much longer and more laborious work, the result of which will be as high as the talent of the artist, but in addition the motion might not be credible: as real characters in the common life surround us, we immediately notice each artifact in the motion of a virtual character.

Several techniques have been developed to provide the animator a way to create a realistic motion using procedural techniques, while letting him a control on the motion [2][3]. Approaches based on space-time optimization have given interesting results in this domain, either to produce full animations 'from scratch' [4][5] or to adapt previously created motion to new environments or characters [6][7]. Using a set of constraints and an objective function, the system generates the best motion with regard to the objective function that respects the constraints.

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 135–144, 2001.

Concerning the simulation of the behavior, procedural methods are generally used when dealing with real-time applications, or that the number of entities is too big to animate them one by one (crowd animation). Techniques using the artificial life tools, such as neuronal networks or learning classifier systems produced some convincing results.

We introduce in this paper a new pipeline dedicated to real-time character animation, from the behavioral simulation to the geometrical aspect. A behavioral engine is used to generate the high level orders, while a dedicated system computes the final motion. Two problems make the jointly use of these both techniques in the same real-time animation pipeline difficult: the generation of some constraints and an objective function on the fly from the high level orders given by the behavior system, and the use of an optimization algorithm in real time. Based on a multi layered architecture, the described pipeline intends to solve these two problems using on the one hand a set of hierarchical scripts to handle the interaction between the behavioral simulation and the animation engine, and in the other hand a method based on optimization to generate the final motion, largely adapted to the real-time problems. We implemented this pipeline in a real-time animation system called LIVE (Life In Virtual Environment).

The second part describes how the system is organized, and how the different sub-systems interact the ones with the others. In the third part, the script system is explained. We show how hierarchical scripts can be used as well to write scenarios (higher level of abstraction) as to control finely the motion at the geometrical level, and how we use them in our system. The low-level motion generation is detailed in the fourth part. A first method based on motion blending is briefly presented. Another one based on spacetime optimization is described more in details. Finally, we present in the last part two applications that use our system to control the motion of characters in real-time environments.

2 System Overview

LIVE is an animation system that works on articulated rigid bodies, organized in a hierarchical bones structure, called entities. High-level orders can be provided to the entities by a behavior engine or by direct orders from the final user. That way, the system can be used to tell interactive virtual stories. High-level orders are compiled on the fly onto a set of low-level tasks using a multi layered script language. The task is the elementary brick of the animation. Each task is associated with one motion generator (or MCM) and is applied to a set of bones. These tasks are the input parameters to one of the two animation techniques described above: they can serve as description to generate motion generators, or can be translated into a set of constraints and an objective function for use with a real-time space-time optimizer. In this last case, orders must be supplied with anticipation to the animation system, as the optimizer works on incoming time segments.

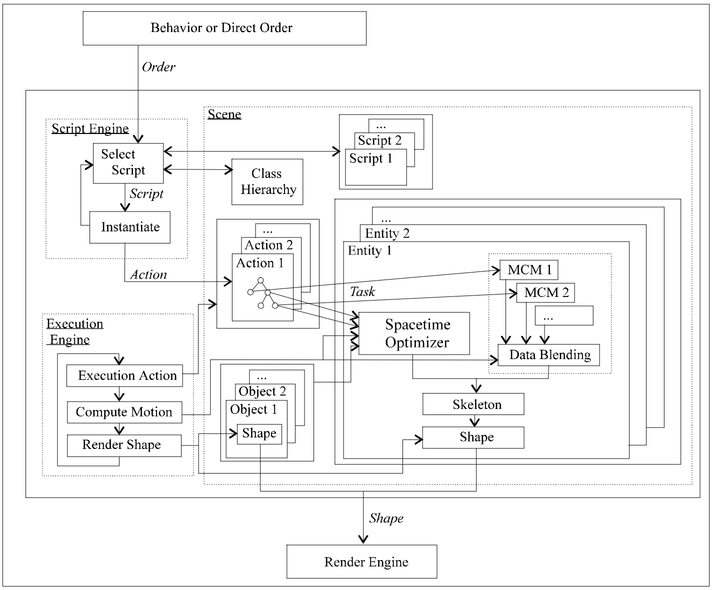


Fig. 1. LIVE architecture

3 Script Language

Several systems use a script language. Perlin developed a human animation system Improv [8] including scripting language to provide a sufficiently powerful interface for instructing virtual humans. In the ACE project [9], agents are controlled by a Lisp behavioral system. A project in the University of Pennsylvania includes: a low-level motor skills with Jack [10], a mid-level parallel automata controller [11], a high-level conceptual representation for driving humans through complex tasks with a Parameterized Action Representation (PAR) [12] and an expressive motion engine (EMOTE) [13]. In the same team, Levison [14] developed architecture (OSR) for an intermediate-planning module that tailors high-level plans to the specific needs of the agents and objects.

In our architecture, the scripts provide an interface between the behavior system and the low-level animation system. The use of multi-layered scripts allows the specification of motion at each level of detail (from the highest semantic order to the geometrical aspect).

Each script defines an action, and is composed of two parts, the heading and the body. The heading identifies the action and holds the list of parameters. The body describes how to perform the action. The elementary command in the script is the brik. There are three kinds of briks: task brik, control brik and call brik. The task brik creates a task associated with one motion generator. The control brik manages the execution of the script. It can also be a control structure (IF, WHILE, …). The call brik uses an action already defined in another script. In each script, all the bricks are executed in parallel. The activation and termination of each brick is controlled by a start and end predicate. Figure 2 shows a part of a walker script.

```
WALK
2
[ENTITY] entity human
[VECTOR3D] position
[ACTION]
...
 [WHILE] loop END_BRIK(first_step)
SUPERIOR(DISTANCEXY(POSITION(entity),position),BD(entity,2_step))
  /* left step */
  [BRIKMOCAP] left_step TRUE END_CAPTURE
    entity LEGS left_step DIRECTIONXY(DIRECTION(POSITION(entity),position))
NEXT [END]

  /* right step */
  [BRIKMOCAP] right_step END_BRIK(left_step) END_CAPTURE
    entity LEGS right_step DIRECTIONXY(DIRECTION(POSITION(entity),position))
NEXT [END]

  /* swing right arm */
  [BRIKKEY] swing_right_arm_behind BEGIN_BRIK(right_step)
END_BRIK(right_step)
    entity RIGHT_ARM BONE (entity,RightArm,VECTOR(0,-2,0))
BD(entity,speed_arm) INF(LOW,1,1) [END]
.... 
 [ENDWHILE]
...
```

Fig. 2. Walker script

The use of call brik allows generating scripts of higher level. This type of scripts uses scripts of lower level to realize a more complex action or to execute a scenario. The scenarios are going to appear as a succession of orders (elementary action) written in natural language. The scripts of higher level (scenarios) generally don't have any parameter. They appear as a list of call brik. Writing a scenario is very simple, it's enough to define every action in natural language with a start and an end predicate. Predicate can be either duration, or an event such as the beginning or the end of another brik. It can also be a geometric information (inverse kinematics, for example, will stop when the end effector reaches the target). The execution of several briks can be chained in a sequential order, in which case the system waits for the end of a brik to start the execution of the next one. That way, actions can be executed in sequence or in parallel.

The scene contains a list of scripts describing all the possible actions. When an order arrives, the system must find in this list the corresponding script. For that purpose, we use a selection engine that filters all the scripts by using a class system. Each entity belongs to a class, and a hierarchy determines the relationship between the different classes. An entity can use only the scripts corresponding to its class and to its ancestor classes.

The selection engine uses three successive filters. First, it selects the script with the same name and the same number of parameter that the order. Secondly, it uses the class hierarchy to find the right script. For each parameter of the entity type, the system tests if the order parameter's class inherits the script parameter's class. If this is the case, the system computes the distance between the two classes in the inheritance tree. When all scripts have been tested, the system selects the script with the smallest distance.

When the script is selected, the system transforms it into an action. An action is an instantiated script. The system creates a copy of the script and replaces each script parameter by the actual scene data. Each action produce in real-time a set of tasks to be executed. These tasks serve as input parameters for the animation system. User can choose between one of the two animation systems that coexist in our system.

4 Low-Level Motion Generation

4.1 Data Blending

When several tasks are executed in the same time on a same bone, the different motions must be blended [3]. We use for that an extra layer inserted between the MCM library and the renderer called blending layer (see fig 1). It allows the motion generators to be used co-operatively or concurrently. The blending layer receives a set of orientations for each bone and computes the final orientation. Extra parameters are needed to give each task a priority (high, medium, low) and a weight. Weights can be animated over time to allow for smooth transitions between several sequential tasks. A complete description of this layer can be found in [2].

4.2 Optimization

The second animation system uses a spacetime optimization to generate the motion. This technique has already been widely used to produce offline animation, but extra processing has to be made before its use in a real-time context.

Background and Related Work. As opposed to other fields, such as rendering or animation of inanimate objects, direct simulation cannot be used to animate characters. To be able to use such a method, it would be necessary to be able of feigning the muscular contraction of all the muscles of a character to get all the internal forces involved in the motion. Given that such a technique is not easy to achieve, other methods have been developed. Whether they use empirical knowledge (such as keyframe animation), data capture from real world, or procedural techniques (inverse kinematics, forward dynamics using empirical data or procedural algorithms to generate the internal forces (activation functions [15]), balance controller [16], non-penetration controller…) none of them is fully satisfactory. As mentioned above, the more empirical the method is, better the control of the animator is. To combine the advantages of each of these methods, techniques have been developed to mix them. The same problem arises for all these methods: first of all, as each method works on different parameters (forces or rotation for example), the mixing can be applied only at the lowest level (the geometry). Secondly, as each method is independent from the others, mixing the result of each of them at the geometrical level leads to conflicts, that must be resolved using for example a priority or a weighted mean algorithm. If only one method is chosen, the benefits of the others are lost. If several methods are blended, the benefits of each of them may be lost. The ideal algorithm would choose the best of every method while taking into account constraints of all the others. This is by essence an optimization problem: from the structure of the character, the environment (which generates implicit constraints, such as the not penetration of the foot in the ground), and control methods we wish to apply, we want to find the best motion (according to a criterion to be defined) which respects all the constraints. The original spacetime constraint method is exposed in [4]. Solving this problem in both the spatial and temporal domains is much more powerful than solving it locally (in time), as the motion to be made or the constraint existing at the time t+1 may influence the motion at time t.

Several specific problems appear when using this kind of methods in a real-time context. As opposed to the former method, input parameters are not a set of tasks. Before the optimization in itself can be done, an objective function and a set of constraints must be generated from the set of control methods and the environment. This extra processing has to be done in real-time, and can not thus be guided by an animator as in an offline context. A second problem lies in the use of an optimization algorithm in real-time.

Motion Representation. To represent the motion over a time period, we choose to use B-Splines bases. For each degree of freedom (DOF), a 1D cubic B-Spline gives the value of the parameter (either a translation value for the root bone, or an angle value for a rotational bone) with respect to time. B-Splines are defined by a set of control points. The advantages of choosing such a representation are multiple. First, cubic B-Spline ensures C^2 continuity. Second, a limited number of variables (the control points) are needed to represent the variation of the parameter over time. The more control points we use, the more accurate the result will be, but the longer the computation time. We can also choose to use non-uniform B-Splines, in which case the knots vector must also be computed. Finally, as control points have a local influence over the curve, only a few of them must be warped to make the curve have a given value at a given time. This is important in an optimization context.

Mathematically, if we choose to use c control points, the motion over the optimized segment of time for the i^{th} DOF $\theta(i)$ is given by:

$$Q_i(t) = \Sigma_{j=0\ldots c} P_{i,j} N_{j,4}(t) . \qquad (1)$$

where $P_{i,j}$ is the j^{th} control point of the i^{th} DOF.

If the knots vector is optimized too (except the 4 first and last value, that are fixed to make the spline pass by the first and last control points), there are $(2\times c-4)$ parameters to optimize for each spline.

Let $P = (p_0, p_1,\ldots,p_{np})$ be the vector of parameters to optimize. This vector defines the entire motion of the character over the time segment. For a skeleton having n DOF, the number of parameters np is given by:

$$np = n \times (2\times c - 4) \text{ if knots vector is optimized} \qquad (2)$$

$$np = n \times c \text{ if knots vector is not optimized}$$

Experience shows that optimizing the knots vector increases a lot the computation cost. It's a better idea to increase the size of the optimized segment of time. The size of this buffer greatly influences the result of the final motion. If the optimization is done on a too small duration, the result will not be interesting, as all the benefit of spacetime optimization will be lost. If this size is too high, real-time cannot be achieved anymore, as the optimization problem cannot be solved sufficiently fast.

Constraints and Objective Function Generation. In order to achieve real-time animation, simplifications have to be made to the original method exposed in [4]. The main difference is that instead of generating movements from scratch, we use one or more original motions and adapt them to compute the final solution. This method is the one used by Popovic [7] and Gleicher [6], and has already given interesting results. These source motions are derived from the motion capture, direct kinematics (from procedural techniques), and dynamics tasks. In these last two cases, the motion is first executed in a separate buffer and stored as a set of B-Splines. The motion captures are also converted to B-Splines. These types of tasks do not generate constraints. They can serve either as initial guess or as new terms in the objective function. This provides an efficient way to ensure that the generated motion will not differ too much from the original motions, and let a great control to the artist. As the source motions are mainly captures and can thus be considered as physically correct, the 'respect Newton's laws' constraint is not included into the set of constraints. This also enables to let intact the creativity of an artist, who can create non-realistic animations, and greatly speeds up the solving.

Each inverse kinematics task generates a constraint. Note that inverse kinematics is also an optimization problem. It can be solved locally using spatial optimization like in traditional methods (for example at each frame), but solving the inverse kinematics task in the spacetime optimization is much more powerful, as it prevents the problems of discontinuity of the motion that can appear with frame per frame methods. Other constraints are provided by the entity's database (mainly joint's angles limitations) and the environment (non-penetration with ground constraints for example). Finally, additional implicit constraints can also exist, such as the constraints on the knots vector.

The objective function to minimize can be simply a measurement of the difference between the computed motion and the original captures (or the captures provided by dynamics simulation or direct kinematics). This function can also serve for giving some characteristics of the motion: minimization of the energy consumption, balance controller by minimizing the distance between the center of mass and the polygon of sustentation, kinematics smoothness, etc.

Let f(P) be the objective function, and C = { C_0(P), C_1(P),…, C_{nc}(P) } be the set of constraints. Constraints can be either equalities (typically inverse kinematics constraints) or inequalities (non-penetration or angle limitations for example).

The problem to solve is now:

minimize f(P)

subject to

$\mathbf{C_i(P) \leq 0}$ i = 0…k

$\mathbf{C_i(P) = 0}$ i = k+1…nc

Solving the Optimization Problem on the Anticipation Buffer.
Once the spacetime optimization problem is fully specified, it can be solved using one of the existing minimization methods. Many different techniques exist, that differ from the form of the function to minimize (linear or quadratic for example), whether or not the derivatives are provided, whether it is a constrained or unconstrained problem, and if necessary the form of the constraints. An alternative method can be to transform the constraints into soft constraints, which are added to the objective function. However, such a technique is difficult to parametrize, as weight must be assigned to constraints. We choose to use an optimizer based on the Sequential Quadratic Programming method. We use the FSQP library [17], which solves large-scale constrained non-linear minimization problems.

System Overview. When using spacetime optimization, the entire animation pipeline must run with anticipation. Orders from the behavioral system or the final user are processed as soon as they arrive to the animation system. A new optimization sequence is run at regular intervals in a separate thread from the main one. Tasks are generated before their execution, and translated on the fly into a set of constraints and an objective function, and an initial guess is computed separately for the parameters of each DOF. After being initialized, the optimizer is then run (multiple instances of the optimizer can run at the same moment). When one optimizer converges to a solution, or when the current time enters the interval where the optimizer runs, it is pushed onto a list of playback optimizers.

Motion playback. The list of already optimized motions is used to playback the final motions in the main thread of the character. A weighted average of the motion computed by each optimizer for the current time is used, with higher weights for the most recent results.

5 Results

We implemented all of these concepts in LIVE, and experiments were made using a behavior engine based on a classifiers system. A first application is dedicated to the simulation of a character walking between several targets on an embossed terrain. The behavior engine can run on a separate machine, and communicates with the script engine via a distributed system. It outputs high-level actions (the targets to reach) on the entities in real-time, and the blending layer is used to compute the final motion. The animation system runs on a common PC-based workstation. The second experiment uses the spacetime optimizer to animate the motion of a single creature. The animation of the creature is based on simple source motions such as walking or running, and aims to adapt these motions to the new situation. Due to the computation times, the time has to be scaled to let the solver find a reasonable solution at each segment of time. Depending on the number of control points and the constraints, a 10 to 100 factor is used.

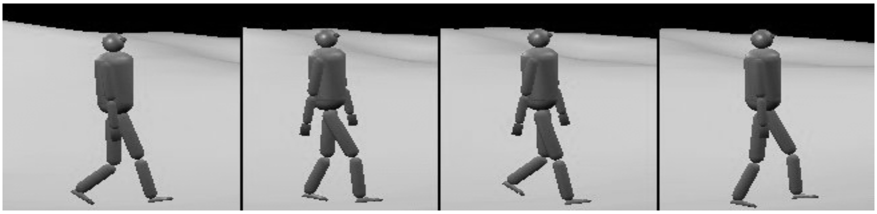


Fig. 3. Real-time animation of an autonomous entity

6 Conclusion and Future Work

A new system has been presented, that uses both a multi layered script system and two animation systems to achieve a convincing realism in real-time character animation, from the behavioral aspect to the generation of good-looking motion. Depending on the context and the requirements, one can choose between one of the two animation systems to compute the final motion. Experiments shows that whereas convincing real-time simulation is possible using a blending method, several concessions have to be done before the use of spacetime optimization is available in real-time. Future work will focus on the improvement of the optimization solver, as well as on a reactive system of constraints generation that responds in real-time to situation changes. Finally, mixed techniques could be considered for the low-level animation subsystem, to automatically restrict the use of spacetime optimization to situations where the blending method fails to achieve sufficiently good results.

References

1. Zicheng Liu, Michael F. Cohen: Keyframe Motion Optimization by Relaxing Speed and Timing, 6th EuroGraphics Workshop on Animation and Simulation (Maastricht, The Netherlands) (1995)
2. Bonnafous, V; Menou, E; Jessel, J.P; Caubet, R: Co-Operative and Concurrent Blending Motion Generators, WSCG (2001)
3. Boulic, R; Becheiraz, P; Emering, L; Thalmann, D: Integration of Motion Control Techniques for Virtual Human and Avatar Real-Time Animation in Proceedings of the VRST'97 (1997) 111-118
4. Andrew Witkin, Michael Kass: Spacetime Constraints, John Dill, editor, Computer Graphics (SIGGRAPH '88 Proceedings), volume 22 (1988) 159-168
5. Michael F. Cohen: Interactive spacetime control for animation, Computer Graphics, 26(2) (1992) 293-302
6. Michael Gleicher, Peter Litwinowicz: Constraint-based motion adaptation, The Journal of Visualization and Computer Animation, 9(2) (1998) 65-94
7. Zoran Popovic, Andrew Witkin: Physically Based Motion Transformation, Computer Graphics Proceedings. ACM SIGGRAPH (1999)
8. Perlin, K, Goldberg, A: Improv: A System for Scripting Interactive Actors in Virtual Worlds, Computer Graphics (1996) 205-216

9. Kallmann , M; Monzani, J.S; Caicedo, A; Thalmann, D: ACE: A Plateform for the Real Time Simulation of Virtual Human Agents, EGCAS'2000, 11th Eurographics Workshop on Animation and Simulation (2000)
10. Badler, N; Palmer, S.P; Bindiganavale, R: Animation Control for Real-Time Virtual Humans, Communication of the ACM, Vol.42, No.8 (1999) 65-73
11. Granieri, J.; Becket, W.; Reich, B.D: Behavioral control for real-time simulated human agents, Symposium on Interactive 3D Graphics (1995) 173-180
12. Badler, N; Bindiganavale, R; Bourne, J; Palmer, M; Shi, J; Schuler, W.: A Parameterized Action Representation for Virtual Human Agents, Workshop on embodied conversational characters, Lake Tahoe, California (1998)
13. Zhao, L; Costa, M, Badler, N: Interpreting Movement Manner, Procedings of Computer Animation 2000 (2000)
14. Levison, L; Balder, N: How Animated Agents Perform Task: Connecting Planning and Manipulation Through Object-Specific Reasoning, Presented at the AAAI Spring Symposium : Toward Physical Interaction and Manipulation (1994)
15. X. Tu, D. Terzopoulos: Artificial Fishes: Physics, Locomotion, Perception, Behavior, Proc. Siggraph 94, ACM Press, New York (1994) 43-50
16. Badler, N; Philips, C.B; Webber, B.L.: Simulating humans, Oxford University Press (1993)
17. AEM Design, CFSQP Homepage: http://www.gachinese.com/aemdesign/FSQPframe.htm

Characters in Search of an Author: AI-Based Virtual Storytelling

Marc Cavazza, Fred Charles, and Steven J. Mead

School of Computing and Mathematics, University of Teesside,
UK-TS1 3BA, Middlesbrough, United Kingdom
{m.o.cavazza, f.charles, steven.j.mead}@tees.ac.uk

Abstract. In this paper, we present the first results obtained with an interactive storytelling prototype. Our main objective is to develop flexible character-based systems, which nevertheless rely on narrative formalisms and representations. Characters' behaviours are generated from plan-based representations, whose content is derived from narrative formalisms. We suggest that search based planning can satisfy the real-time requirements of interactive storytelling, while still being compatible with the narrative formalisation we are pursuing. We then describe into greater detail a short episode generated by the system, which illustrates both high-level results and technical aspects, such as re-planning and user intervention. Further work will be dedicated to developing more complex narrative representations and investigating the relations between natural language semantics and narrative structures in the context of interactive storytelling.

E dov'è il copione?
— È in noi, signore.

Luigi Pirandello, *Sei personnagi in cerca d'autore*

1 Introduction

In this paper, we describe the principles behind a virtual interactive storytelling prototype, in which a generic storyline played by artificial actors can be influenced by user intervention. The final applications we are addressing consist in being able to alter the ending of stories that have an otherwise well-defined narrative structure, thus reconciling interactivity with story authoring. Ideally, this would make possible to alter the otherwise dramatic ending of a classical play towards a merrier conclusion.

There has been much recent work in interactive storytelling that has developed a wide range of perspectives; emergent storytelling [1] [2], user-centred plot resolution [3], character-based approaches [4] [5], anytime interaction [6] and the role of narrative formalisms [7]. This work has identified relevant dimensions and key problems for the implementation of interactive storytelling, among which: the status of the user, the level of explicit narrative representation and narrative control, the modes of user intervention, and, most importantly, the relations between characters and plot.

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 145–154, 2001.

Some of the above problems derive from the inherent tension between interaction and narrative [4] [8]. Interactive systems demand user involvement but often at the expense of a real storyline; on the other hand, a strong narrative dimension is traditionally conceived of with a user as spectator rather than actively involved. Our own solution to this problem consists (in accordance with our final objectives stated above) in limiting the user involvement in the story, though interaction should be allowed at anytime. This is achieved by driving the plot with autonomous characters' behaviours, and allowing the user to interfere with the characters' plans. The user can interact either by physical intervention on the set or by passing information to the actors (e.g., through speech input). In this context, the most important aspect of interactive storytelling is the relation between characters and plot. In his now classic play, Pirandello imagined that characters could be collectively in possession of the plot [9]. This is the best illustration, in modern times, of the duality between character and plot, much debated since its introduction by Aristotle [10]. In the next sections, we develop the hypothesis that narrative functions describing a story can be used to generate plan-based behaviours for the characters. Further, we propose that the respective roles for the various characters of a story should be defined from high-level narrative principles, in a similar fashion.

A first interactive storytelling prototype has been fully implemented and runs in a real-time interactive 3D environment [11]. Graphic rendering, character animation and user interaction in the system are based on the Unreal Tournament™ game engine. This engine provides high-quality display at a constant frame rate, while also serving as a software development environment [12]. Besides embedding its own scripting language (UnrealScript™), it can integrate complete C++ modules or communicate via sockets with external software modules. This prototype has been developed in C++ and UnrealScript™, and runs a simple scenario that we describe in the next sections, together with some of the results obtained.

2 Characters-Driven Storytelling and Narrative Formalisms

The storyline for our experiments is based on a simple sitcom-like scenario, where the main character ("Ross") wants to invite the female character ("Rachel") out on a date. This scenario comprises a principal narrative element (will he succeed?) as well as situational elements (the actual episodes of this overall plan that can have dramatic significance, e.g., how he will manage to talk to her in private if she is busy, etc.). For this reason, sitcoms appeared as an interesting narrative genre to investigate. Our system is driven by characters' behaviours, represented as plans.

The behaviour of our artificial characters is based on AI planning techniques, as introduced by Webber et al. [13] for high-level behaviours of virtual actors and by Young in the specific case of storytelling [4]. Our perspective on the contents of the character's plans is clearly narrative rather than cognitive.

We are endeavouring to define a proper representational content for narratives within the implementation framework of AI planning representations. This objective can be illustrated by analogy with computational linguistics: linguistic formalisms can be used to analyse natural language through parsing, and the same syntactic descriptions can serve to generate sentences and text. In a similar fashion, we suggest that a true "computational narratology" could be created using the formalisms developed by narrative analysis. These formalisms would serve as a basis for narrative representations from which stories would be generated. Let us consider first that the initial step in formalizing a plan is to describe a Hierarchical Task Network (HTN), i.e. a hierarchy of sub-goals and actions corresponding to the overall plan. HTNs will thus be our target representations for characters' plans: section 3 will describe the actual generation of behaviours from HTNs.

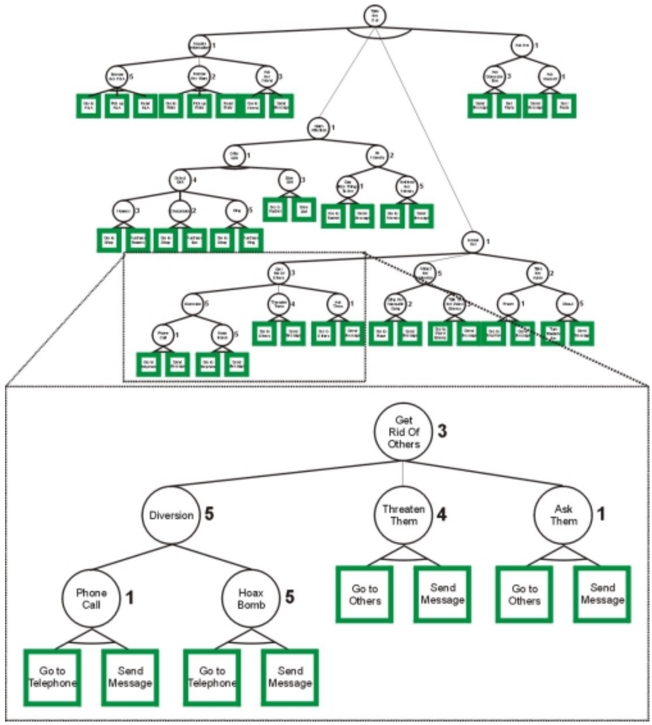


Fig. 1. Ross' Plan

Most work in interactive storytelling has made some reference to narratology. Mateas [8] proposed a neo-Aristotelian framework for interactive stories, Prada et al. [14] referred to Propp's narrative functions as a relevance model for her story-like situations, Szilas [7] has advocated explicit narrative representations based on more recent narrative theories, like those of structuralist authors such as Greimas and Bremond, who have extended narrative formalisms beyond Propp's functions.

A natural approach is to investigate the kind of formalisation attempted in narratology and to determine whether it can be made more computational. Indeed, two sorts of tree-based representations have been introduced in structural narratology, mainly by Barthes: the stemma-like representation [15] and the *proairetic* tree [16]. The former illustrates the decomposition of narrative functions into temporal sequences of lower-level functions. This can be illustrated by considering the overall plan for the Ross character (Figure 1). In order to invite Rachel, he must for instance acquire information on her preferences, befriend her, find a way to talk to her in private, and finally formulate his request (or having someone acting on his behalf, etc.). The latter makes explicit the choices that a character can make at various points[1], i.e. the alternative actions he can take. Dynamic choice of an actual course of action is actually the basis for plot instantiation in interactive storytelling, as otherwise suggested by Young [4]. A given plot will correspond to one and one choice only, together with its long-term consequences. In our scenario, Ross can choose to isolate Rachel from her friends by attracting her attention or by rudely interrupt the conversation. These options (among others) have obviously quite different consequences.

To summarise, we can say that a narrative representation would be an HTN, whose nodes are constituted by various levels of narrative functions, the relationships between the various levels representing composition or alternatives. The HTN represents more than the "role" of the character, as it encompasses potential all variants, at every level. The final level of the HTN consists in terminal actions, i.e. those actions actually played by the character on the virtual stage. A narrative function can correspond to sub-goals at different levels of hierarchy (acquire-information, isolate-her, offer-a-gift): the important point is that the predicative structure of narrative functions (i.e., the other actors involved) is deferred to the lowest compatible level of the hierarchy. This predicative structure is also a function of the main character from which perspective the HTN is described. The initial storyline should actually determine not only the main character plan, but those of other characters as well. This separate definition of roles shall serve as a basis for the dynamic generation of story variants, as individual characters' plans will interfere with one another, depending on initial conditions and pseudo-random factors. The problem of dependencies between characters' roles has been described within modern narratology, though not to a formal level. Narrative functions can be refined into bipolar relations between couple of actors, emphasising the asymmetry in their roles [15]. We have adopted this framework to define the respective behaviours of our two leading characters. We started with the overall narrative properties imposed by the story genre; sitcoms offer a light perspective on the difficulties of romance: the female character is often not aware of the feelings of the male character. In terms of behaviour definition, this amount to defining an "active" plan for the Ross character (oriented towards inviting Rachel) and a generic pattern of behaviour for Rachel (her day-to-day activities, subject to variations according to her mood and feelings). This is illustrated on Figure 2. There is significant evidence from narratology studies in favour of this approach: what we have described here is very similar to the respective roles of the main characters in Balzac's novel *Sarrasine*, which has been entirely analysed by Barthes in his seminal book, "S/Z" [16].

We can now propose a very preliminary methodology for the definition of roles within an overall storyline:

[1] Barthes uses the concept of *proairesis*, or choice between various courses of action, with reference to Aristotle [16].

- identify the various roles and the main feature characters
- describe the roles for these main characters as generic plans. In doing so, the predicative structure of narrative functions is refined: narrative functions incorporate reference to other characters in their definition: ask-*her*-out(Rachel), ask-*her*-friend(Pheobe), etc.
- enhance the role of feature characters with *proairetic* variations at the appropriate level of description

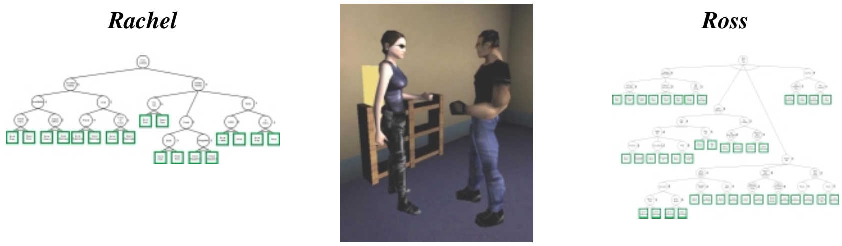


Fig. 2. Ross' "active" plan and Rachel's generic pattern of behaviours

Another important concept in interactive storytelling is causality [10], as it supports the consequences of interaction, whether it be agent-agent interaction or user intervention. Some interactive storytelling systems make causality explicit in their representations, for instance by using an ATMS [3]. However, in a task network representation based on actions and sub-goals, causality is not explicitly represented. One form of implicit causality is the enabling of further actions by their predecessors in the HTN ordering, but it is not related to interaction and dynamic generation. Other forms of causality are implicit as well: for instance, if, when attempting to talk to Rachel in private, Ross behaves rudely with Pheobe, he might actually upset Rachel and cause her to change feelings in his regard (see the example of section 4). This point illustrates an important practical equivalence between choice and causality, which has been described by Raskin [10]. In a plot-based approach [3] causality can be explicitly represented: in a character based approach, the proairetic aspect is dominant. The character-plot duality has thus a translation in terms of causal representations.

3 AI Planning for Characters' Behaviour

AI planning is used to implement characters' behaviour. The planning mechanism should produce a real-time plan from the plan-based narrative representations. An essential requirement, common to all virtual actors evolving in dynamic environments, is that planning and execution should be interleaved [17] [18]. A specific constraint of interactive storytelling is that actions executed by the characters should be properly played in the context of the story: we call this aspect, which concerns the visual presentation of action, *dramatisation*. This is an important aspect, as the user will determine his intervention (if any), according to the meaning he attributes to the characters' actions. Finally, the approach should support re-planning when the initial plan fails, due to interference from other characters or the user.

The planning system generates a plan for each character, using the HTN describing its own role. From a formal perspective, if we assume that the various sub-goals are independent, planning can be directly achieved by searching the AND/OR graph corresponding to the HTN [19]. In the generic case, the HTN would have to undergo a complex linearisation process beforehand, but in the case of goal independence the solution plan is a direct sub-graph of the HTN[2]. The system can thus use an algorithm such as AO* to produce a solution sub-graph, whose terminal nodes form a sequence of actions to be played by the virtual actor [11]. The standard AO* algorithm comprises a forward expansion, which generates a solution basis (the best solution sub-graph), and a backwards propagation from terminal nodes, which updates the value of the solution expanded. However, AO* alone does not meet our requirements for planning. We have thus developed a "real-time" variant of AO* that interleaves planning and execution. Our planner uses left-to-right depth-first search in a similar fashion than the MinMin algorithm of Pemberton and Korf [20]. It plans forward, until it reaches the first activable terminal action[3], which is then carried out by the character and appropriately dramatised in the virtual environment. The outcome of action execution can be propagated back into the search process by taking advantage of the roll-back mechanism which is part of AO*, which is essentially deferred to action execution in our real-time variant. This supports re-planning on action failure, which is one of the mechanisms for story generation. For instance, Ross wanted to know Rachel better by reading her diary. But the user has hidden the diary in a safe place (Figure 3). Ross started to execute his plan, but only realised the diary was missing when reaching its default location. Ross has to find a new way of acquiring information on Rachel, and re-plans a solution for that sub-goal, which consists in asking her friend Phoebe. He starts the execution of this new sequence by looking for Phoebe.

Top-down planning alone is not sufficient to cope with the executability conditions of actions in a dynamic environment. For instance, Ross might not want to steal Rachel's diary if he can be spotted by Monica. To solve this kind of problem, Geib and Webber have proposed to complement top-down planning with situated reasoning [21] [22]. This can be successfully applied in conjunction with real-time planning: the main plan can be interrupted to cope with situated actions and control later returned to the plan after updating the action pre-conditions. For instance, if at an early stage of his plan, Ross bumps into Rachel in a corridor, he cannot just ignore her, but this situation cannot be incorporated into the top-down plan: it has to be treated specifically. Situated reasoning can take place at various levels of the plan hierarchy. For instance, it can enforce generic "priority" behaviours, like avoiding Rachel at the early stages of the plan. At the lowest level of the plan, situated reasoning can even constitute an alternative to re-planning. For instance, if Ross wants to read Rachel's diary but she is

---

[2] To a large extent, sub-goal independence appears to be a property of narrative representations, though this point deserves further investigation.

[3] A similar approach, has been described by Geib [21] as part of his incremental planner "ItPlans". However, our system proceeds depth-first towards the first executable action.

using it, he can wait for her to finish (unnoticed by her), rather than find another source of information. This form of situated reasoning is based on the duration of actions and the nature of their resources.

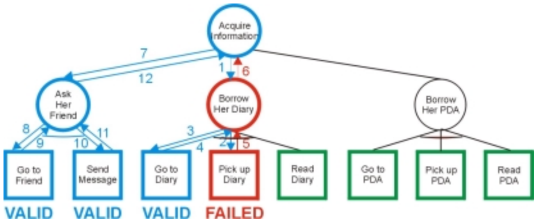


Fig. 3. Re-planning following user intervention

The final story instantiation is mostly determined by the interaction between characters, i.e. by how their plans eventually result in joint action. At the algorithmic level (i.e., RTAO*), there are no intrinsic synchronisation mechanisms between the search processes that drive the two character's behaviours. Rather, using similar principles to those governing user intervention, the characters can interact via their physical environment, by competing for resources for action (narrative objects or other characters). For instance, Ross can influence Rachel's activity and make her more available by e.g., taking care of one of her duties without telling her. Or he can look for some information from Phoebe, but she left to do some shopping with Monica. It is this interaction between the two behaviours that produces much of the situational elements of the story (apart from the final conclusion). To some extent the actual plot can be seen as the "cross product" of the characters plans[4]. There is a number of factors that concurs to make the plot not predictable from the user's perspective, mostly related to actions' duration and competition for action resources. For instance, depending on their initial random positions, some actors can engage in a conversation and become unavailable to others. Similarly, they can be first to reach some objects of narrative importance (telephone, diary, etc.), which will cause other characters' to change their plans, creating new interactions, etc. However, the important point is that characters always keep track of their long-term goals, which differentiates interactive storytelling from simulation-based computer games.

Finally, user intervention is another source of plan variability. The user can interfere with either the execution of the plans' terminal actions or with the plan goals: this determines two major modes of interaction: physical interaction and linguistic interaction [23]. Physical interaction takes place when the user interferes with plans resources on the virtual set, for instance by stealing an object that the artificial actor might use to reach its goal (the "diary" example above). Linguistic interaction is based on speech input that directly passes information to the artificial actors, altering their intentions and goals. For instance, the user can issue a recommendation such as "try to be nice", using speech recognition. This should rule out any rude behaviour towards Rachel or her friends, such as sending her friends away to talk to her. Once again, the

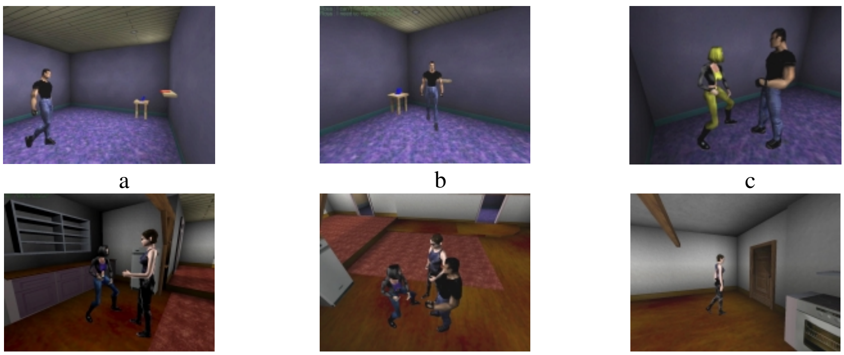
[4] This interesting metaphor was suggested to us by an anonymous reviewer.

effects of such a "doctrine statement" [13] can be implemented by revising the heuristic values attached to certain node categories in the plan graph.

4 Dynamic Story Generation: First Results

In this section, we describe into further detail a complete example obtained from the prototype. This example will illustrate user intervention, re-planning and the use of moods to propagate causality between various characters.

In order to get the information he needs, Ross goes to read Rachel's diary (a). However, he realises that somebody moved the diary (b). So instead, he decides to meet Phoebe to ask her about Rachel (c). In the meantime, Rachel is talking to Monica (d). In order to talk privately with Rachel, Ross is ordering Monica to leave (e). Rachel gets upset and ostensibly leaves the room (f).



a b c

d e f

Fig. 4. Sequence of actions illustrating the story instantiation

Let us now give a more technical description of these events, by detailing the associated steps in plan generation or terminal actions. Each of the main characters has its own planning system: they are synchronised through Unreal™ low-level mechanisms. Firstly, Ross' plan. The first sub-goal for Ross' plan is to acquire information about Rachel. There are various ways to satisfy this goal in Ross' behaviour representation, and the first one selected is to read her diary. The corresponding script involves going to the diary location and reading it. When Ross arrives in sight of the diary, the pre-condition of the action of "reading it" is checked: the diary is in place. This pre-condition is not satisfied, as the user intervened by removing the object from the set. Hence the second terminal action "ReadDiary" fails, as well as the whole sub-plan. The re-planning produces a new partial solution, which consists in asking Phoebe. Ross then goes to Phoebe's location and starts talking to her. As Phoebe is a reactive actor, she responds directly to Ross' request, in this case positively. In the meantime, Rachel's plan that governs her spontaneous activity, determines her to talk to her friend. She reaches Monica and starts conversing through a durative action (a scripted action which is associated a clock based on the internal Unreal™ clock). When Ross

has finished talking to Phoebe, he needs to isolate Rachel in order to ask her out. The pre-conditions for a terminal action involving conversation with another actor is to check whether this actor is free. The personality profile defined initially for Ross (character with no ruthless manners) influences the heuristic values of the sub-plan nodes. So, Ross interrupts Rachel's conversation and, in a rude way, asks Monica to leave. Rachel reacts consequently to the situation by displaying a relevant mood state: she gets upset. Internally, the mood state is altered accordingly: all heuristics are revised, and of course, the activity "Talk to Monica" fails. Rachel leaves the room. In the same way, Ross' low-level mechanisms will provide situational information that will modify his internal states and influence his sub-plans. Ross will run after her when he realises Rachel is upset.
To summarise, this example illustrates the interaction of the two main character's plans, also influenced by user interference. Though these plans are designed from global narrative principles (considering the story genre), they are run independently. The particular interactions that take place depend on a number of variable factors, which contribute to the diversity of plots generated.

5 Conclusion

We have described an interactive storytelling system, which attempts to reconcile the character-based approach with the use of sophisticated narrative formalisms. At the heart of our system is a specific conception of user interaction. Namely, that the user can alter the ongoing events within the limits of the narrative genre itself. This amounts to saying that the entertaining aspects derive from some form of empowerment of the user, not to be passively linked to the plot: this form of interactivity is also a consequence of our emphasis on narrative structures.

There are many challenges related to this approach, in particular in further exploring narrative theories, with the prospect of implementing complex narratives and multiple storylines. The inclusion of more sophisticated language technologies in interactive storytelling is also a natural long-term goal, especially because of the relations between natural language semantics and narrative structures.

References

1. Dautenhahn, K.: Story-Telling in Virtual Environments, ECAI'98 Workshop on Intelligent Virtual Environments, Brighton, UK (1998).
2. Aylett, R.: Narrative in Virtual Environments: Towards Emergent Narrative. Papers from the AAAI Fall Symp. on Narrative Intelligence, Tech. Rep. FS-99-01, AAAI Press (1999).
3. Sgouros, N.M., Papakonstantinou, G. and Tsanakas, P.: A Framework for Plot Control in Interactive Story Systems, Proceedings AAAI'96, Portland, AAAI Press (1996).
4. Young, R.M.: Creating Interactive Narrative Structures: The Potential for AI Approaches. AAAI Spring Symp. on Artificial Intelligence and Computer Games, AAAI Press (2000).

5. Mateas, M.: An Oz-Centric Review of Interactive Drama and Believable Agents. Technical Report CMU-CS-97-156, Department of Computer Science, Carnegie Mellon University, Pittsburgh, USA (1997).
6. Nakatsu, R. and Tosa, N.: Interactive Movies, In: B. Furht (Ed), Handbook of Internet and Multimedia – Systems and applications, CRC Press and IEEE Press (1999).
7. Szilas, N.: Interactive Drama on Computer: Beyond Linear Narrative. Papers from the AAAI Fall Symp. on Narrative Intelligence, Tech. Rep. FS-99-01, AAAI Press (1999).
8. Mateas, M.: A Neo-Aristotelian Theory of Interactive Drama. Working Notes of the AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment, AAAI Press (2000).
9. Pirandello, L.: Sei personaggi in cerca d'autore (in Italian). Mondadori, Milan (1993).
10. Raskin, R: Five Parameters for Story Design. P.O.V., n. 5 (1998).
11. Cavazza, M. Charles, F., Mead, S.J., and Strachan, A.I.: Virtual Actors' Behaviour for 3D Interactive Storytelling. Proceedings of the Eurographics 2001 Conference (Short Paper), Manchester, UK, to appear (2001).
12. Young, R.M.: An Overview of the Mimesis Architecture: Integrating Intelligent Narrative Control into an Existing Gaming Environment. Working Notes of the AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment, AAAI Press (2001).
13. Webber, B., Badler, N., Di Eugenio, B., Geib, C., Levison, L. and Moore, M.: Instructions, Intentions and Expectations, Artificial Intelligence Journal, 73, pp. 253-269 (1995).
14. Prada, R., Machado, I., and Paiva, A.: Teatrix: Virtual Environments for Story Creation. Proceedings of Intelligent Tutoring Systems 2000, Canada (2000).
15. Barthes, R.: Introduction a l'Analyse Structurale des Récits (in French), Communications, 8, pp. 1-27 (1966).
16. Barthes, R.: S/Z (in French). Editions du Seuil, Paris (1970).
17. Laird, J.E.: It Knows What You're Going To Do: Adding Anticipation to a Quakebot, Working Notes of the AAAI Spring Symp. on Artificial Intelligence and Interactive Entertainment, Tech. Rep. SS-00-02, AAAI Press (2000).
18. Funge, J., Tu, X., and Terzopoulos, D.: Cognitive Modelling: Knowledge, Reasoning and Planning for Intelligent Characters. Proceedings of SIGGRAPH'99, Los Angeles, USA, pp. 29-38 (1999).
19. Tsuneto, R., Nau, D. and Hendler, J.: Plan-Refinement Strategies and Search-Space Size. Proceedings of the European Conference on Planning, pp. 414-426 (1997).
20. Pemberton, J.C. and Korf, R.E.: Incremental Search Algorithms for Real-Time Decision Making. Proceedings of the Second Artificial Intelligence Planning Systems Conference (AIPS-94), pp. 140-145 (1994).
21. Geib, C.: The Intentional Planning System: ItPlans, Proceedings of the 2nd Artificial Intelligence Planning Conference (AIPS-94), pp. 55-64 (1994).
22. Geib, C. and Webber, B.: A consequence of incorporating intentions in means-end planning.Working Notes – AAAI Spring Symposium Series: Foundations of Automatic Planning: The Classical Approach and Beyond. AAAI Press (1993).
23. Charles, F., Mead, S. and Cavazza, M.: User Intervention in Virtual Interactive Storytelling. Proceedings of VRIC 2001, Laval, France (2001).

Virtual Agents' Self-Perception in Story Telling

Eric Maffre, Jacques Tisseau, and Marc Parenthoën

Laboratoire d'Informatique Industrielle (Li2/EA2215)
Ecole Nationale d'Ingénieurs de Brest, CP 15, F-29608 Brest Cedex, France
{emaffre, tisseau, parenthoen}@enib.fr

Abstract. The goal of this work is self-perception modeling of autonomous agents in virtual story telling. It's inspired from work of psychologists and neuro-physiologists. From psychology , we use fuzzy cognitive maps (FCM) to model and implement believable agents' behaviours. These cognitive maps allow us to give not only sensation but also perception, in the sense that our agents perceive environment in function of their inner states or emotions. From neuro-physiology, we implement the idea that movement is simulated in the cortex before it is performed in real world. Virtual agent's self-perception is the ability to simulate different behaviours in its own imaginary space before acting in "real" world. This self-perception implemented by "simulation in the simulation" is one of the keys for the autonomy of virtual entities' decision.

1 Introduction

The goal of this work is implementation of self-perception processus for autonomous virtual agent.

This autonomy is essential for credibility and rests on a sensorimotor autonomy: each entity has sensors and effectors enabling it to be informed and to act on its environment, on an autonomy of execution: the execution controller of each entity is independent of the controllers of the other entities, and on an autonomy of decision: each entity decides according to its own personality (history, intentions, state and perceptions). The virtual human autonomy is one of the current VR stakes, as underlines D. Thalmann in a recent futurology [6].

From the psychologist Tolman [Tolman48], we use cognitive maps to model believable agent's behaviours. Fuzzy Cognitive Maps (FCM) formalism proposed by Kosko [Kosko86] permit to specify and implement character actors. These "charactors" improvise in free interaction within the framework of a "nouvelle vague" scenario, has could Godard do [5].

Then we use neuro-physiologic principle improved by Berthoz: movement is simulated in the cortex before it is performed in real world [1]. The multi-agent environment oRis is used to implement this self-perception [2].

Next section presents FCMs definition and how they can specify and control agent's behaviours. Section 3 describes characters' perception and implementation of self-perception *sensu* Berthoz as a simulation in the simulation, via the interactive story of a montain pasture with shepherd, sheep and dog.

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 155–158, 2001.

2 FCMs and Agent's Behaviour

2.1 FCMs Definition

K is one of the rings $\mathbb{Z}$ or $\mathbb{R}$, δ one of the numbers 0 or 1, $\mathcal{V}$ one of the sets $\{0,1\}$, $\{-1,0,1\}$, or $[-\delta,1]$. Let be $(n,t_0) \in \mathbb{N}^2$ and $k \in \mathbb{R}_+^*$.

A FCM $\mathcal{F}$ is a sextuplet $(\mathcal{C},\mathcal{A},L,A,f_a,\mathcal{R})$, where:

- $\mathcal{C} = \{C_1, \cdots, C_n\}$ is an n concepts set forming the nodes of a graph.
- $\mathcal{A} \subset \mathcal{C} \times \mathcal{C}$ is the set of the arcs (C_i, C_j) directed from C_i to C_j.
- $L : \left| \begin{array}{ll} \mathcal{C}\times\mathcal{C} & \to K \\ (C_i,\ C_j) & \mapsto L_{ij} \end{array} \right.$ is the function from $\mathcal{C}\times\mathcal{C}$ to K associating with a concepts couple $(C_i,\ C_j)$, $L_{ij} = 0$ if $(C_i,\ C_j) \notin \mathcal{A}$, or L_{ij} equals the weight of the directed arc from C_i to C_j if $(C_i,\ C_j) \in \mathcal{A}$. $L(\mathcal{C}\times\mathcal{C}) = (L_{ij}) \in K^{n\times n}$ is an $\mathcal{M}_n(K)$ matrix. It is the FCM $\mathcal{F}$ links matrix which, to simplify, one notes L if it is not ambiguous.
- $A : \left| \begin{array}{ll} \mathcal{C} & \to \mathcal{V}^{\mathbb{N}} \\ C_i & \mapsto\ a_i \end{array} \right.$ is the function associating with each concept C_i its activation degrees sequence such as for $t \in \mathbb{N},\ a_i(t) \in \mathcal{V}$ represents its activation degree at moment t. One will note $a(t) = [(a_i(t))_{i\in[\![1,n]\!]}]^T$ the activations vector at moment t.
- $f_a \in (\mathbb{R}^n)^{\mathbb{N}}$ the extern activations vectors sequence such as for $i \in [\![1,n]\!]$ and $t \geq t_0,\ f_{a_i}(t)$ represents the extern activation of the concept C_i at the moment t.
- $\mathcal{R}$ is a recurrence relation on $t \geq t_0$ translating the dynamic of the FCM $\mathcal{F}$:

$$\begin{gathered} \forall t \geq 0,\ a(t+1) = G\left(f_a(t), L^T \cdot a(t)\right) \\ \forall i \in [\![1,n]\!],\ a_i(0) = 0\,, \forall t \geq 0,\ a_i(t+1) = \sigma(g_i\left(f_{a_i}(t), \sum_{j\in[\![1,n]\!]} L_{ji}a_j(t)\right)) \end{gathered} \tag{1}$$

  where $g_i : \mathbb{R}^2 \to \mathbb{R}$ represents "fuzzy logic" operators between activations from graph of influences $L^T \cdot a(t)$ and extern activations $f_a(t)$, and where $\sigma : \mathbb{R} \to \mathcal{V}$ standardizes activations; usually σ is a sigmoid if $\mathcal{V}$ is a fuzzy set, or crisped sigmoid if $\mathcal{V}$ is a finit set.

2.2 FCMs Relations with Agent's Behaviour

FCMs can specify and control behaviours of virtual agents. They allow specification of believable agents.

Agent has sensors, effectors and decides on its behaviour. FCM in relation with this agent has perceptive and motor concepts. Decision of the agent associated with the FCM is controlled by extern activations and FCM's dynamic. Perceptive concepts are activated by fuzzyfication of sensors, while motor concepts activate effectors by defuzzyfication (Fig.1).

As an exemple, lets specify emotional escape behaviour of an agent with FCMs. One FCM can control flee direction as fuzzy controller could do from angle sensors towards enemy. When another FCM controls agent's speed with fuzzyfication of distance to enemy (Fig.2).

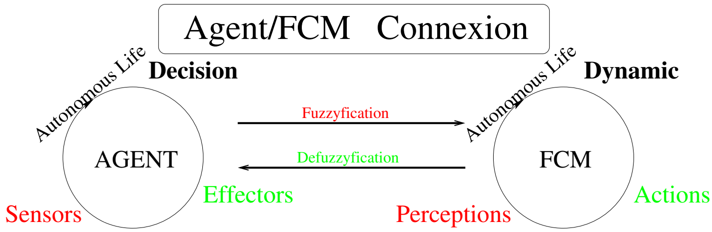


Fig. 1. Autonomous FCM Controls Autonomous Agent Behaviour

3 Perception and Self-Perception

3.1 Virtual Characters' Perception

We distinguish sensation from perception: sensation results from sensors alone; while perception is sensation influenced by internal state. A FCM makes possible to model perception thanks to the links between central concepts and sensitive concepts. Figure 2 illustrates how the fear could modify perception of enemy. Bold connexions from fear concept to perceptive concepts about enemy's proximity specify paranoia character and the auto-recurent one on fear specifies stress.

From prototypic FCMs, we can compute character instances. In fact, we just add weight variations on casual links and obtain different characters.

These ideas are illustrated by interactive story of a mountain pasture with sheep and dog charactors and a shepherd controlled by operator who can give orders to his dogs and walk in his pasture. Orders could be "stop, follow slowly, run, bring back, guard, see far, see close".

3.2 Self-Perception

We have implemented a simulation in the simulation based on neuro-physiologist principle improved by Berthoz: movement is simulated in the cortex before it is performed in real world. The cortex uses projective space deconnected from muscles by inhibitor neurons and anticipates movements. This needs long term

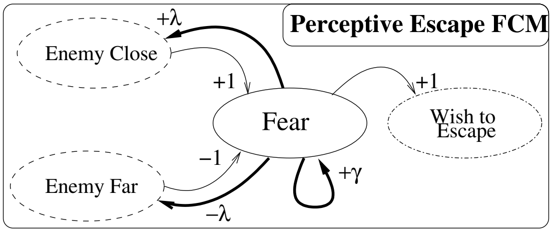


Fig. 2. FCM Specifies Fear and Flee Speed of a Pray in front of Predator

memory for environment's representations and short term memory for instant situation at the moment of simulation. Then result from projective space are use for better adaptation in real world.

We show a dog possessing self-perception: it simulates two different strategies in an imaginary space, then chooses best one for gathering the sheep.

The dog has two vision strategies: everywhere or closed neighbourhood. It simulates these two different behaviours in its simplified imaginary pasture before performing in virtual world. Long term memory corresponds to simplified sheep FCMs while short term memory is determined by observables. This self-perception gives the dog the ability to choose the best behaviour without shepherd's orders.

4 Perspectives

We have seen that FCMs can specify, control and predict charactors' behaviours. These abilities permit implementation of self-perception for believable agents.

This technique opens the gates with a true co-operation between autonomous agents, each one being able to imagine the consequences of strategies. Futhermore, FCMs have learning abilities, which should help in construction of links values via experiments [4]. This self-perception implemented by "simulation in the simulation" is one of the keys for the autonomy of virtual entities' decision.

References

1. Berthoz A., *Le Sens du Mouvement*, Odile Jacob Ed., Paris, 1997.
2. Harrouet F., *oRis: s'Immerger par le Langage pour le Prototypage d'Univers Virtuels à Bases d'Entités Autonomes*, PhD Thesis, Brest University, France, 2000.
3. Kosko B., Fuzzy Cognitive Maps, *International Journal Man-Machine Studies*, 24:65-75, 1986.
4. Kosko B., Hidden patterns in Combined and Adaptative Knowledge Networks, *International Journal of approximate Reasonning*, 2:337-393, 1988.
5. Parenthoën M., Reigner P.,Tisseau J., Dory F., Agent's Perception and Charactors in Virtual Worlds: Put Fuzzy Cognitives Maps To Work, *Proceding of VRIC'01*, 11-18, Laval, 2001.
6. Thalmann D., Challenges for the Research in Virtual Humans, *Workshop Achieving Human-like behavior in interactive animated agent*, Barcelona, Spain, June 3, 2000.
7. Tolman E. C., Cognitive Maps in Rats and Men, *Psychological Review*, 42, 55, 189-208, 1948.

Reflections from a Hobby Horse

Paul Kafno

HD Thames Ltd, 199 Piccadilly, London W1V 9LE, UK.
Hdthames@easynet.co.uk

Abstract. This paper examines ideas that may shape a new approach to story-telling. It considers innovations from antiquity to the present, with examples from the Ad Herrenium, and work by Laurence Sterne, Raymond Roussel, Bernini, Rodin and Picasso and considers their relevance to virtual story creation.

It is one of the fascinations of the modern times, that although we know a great deal about reality, our perceptions are largely shaped by stories. Fiction is everywhere – in music, art, dance, drama, TV, films, radio, computer games, poetry, and, of course, novels. Some of these stories are true – based on real persons and events – others are imaginary. It is perhaps an enigma that imaginary stories that tend to engage us most.

Serious minded people are often worried about this confusion between true matter which they see as "real" and more trivial stuff which they call "fiction". It's an old problem. In the eighteenth century there were folk who would not read fiction because they knew it consisted of lies. Many of us in this room, may think the problem is quite different. We are concerned to create simulations of reality which help us tell stories better and engage our audiences more fully. Yet this question of what is real and what is fictional will not go away, and I will argue that the relationship between the two is fundamental to an understanding of what might happen in the future.

We live in an age of amazing technology. Digital manipulation enables us to control images and build virtual worlds pixel by pixel, polygon by polygon. Signal separated from noise means that we can subtract and combine seamlessly. Artificial intelligence enables us to endow our creations with the illusion of intelligence. Never have creators had so much control. Yet the question is simply – what can we do with it? Is all this stuff merely a replacement? Or can it genuinely lead us towards something new? In this paper I want to look at some of the important ideas that may shape a new creativity, and see that they are not novelties thrown up by technical development, but expressions of a deep human energy that has been bubbling since antiquity.

You may remember that in Shakespeare's Hamlet, the main character, the Prince of Denmark is deeply troubled. Evidence, both real and supernatural has led him to suspect that his uncle has murdered his father. His motive, to get the crown of Denmark and an attractive queen as wife. When a troupe of actors arrive at Elsinore, Hamlet, confident of the power of fiction, persuades them to perform a play which is

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 161–170, 2001.

sufficiently close to the reality of what has happened to flush out the truth. But before that, he is moved by the power of the players' art to ask:

"What's Hecuba to him, or he to Hecuba,
That he should weep for her."

This simple question lies at the heart of the enigmatic power of fiction. It led the philosopher Colin Radford to a serious question. "Fictional characters are not real, and we adults know that. So how can we be moved by what happens to them?" It is an interesting and absolutely fundamental question. What it is to be moved by what happens to other human beings? If someone tells us a tragic story it is most likely that the account will "awaken or reawaken feelings of anger horror, dismay or outrage, and, if you are tenderhearted, you may well be moved to tears. You may even grieve." What then are our feelings on being told that the whole thing is a pack of lies?

In fiction we know from the beginning that things are not true. We become participants in a game. We willingly collude in letting our emotions be brought forth under false pretensions, and we do it because we know, hedonistically, that it makes us feel good. The amazing thing about human beings is that we can be moved in this way, and these acts of sympathy – or empathy – are essential elements in the business that we call art. Indeed we pay money for the books, theatre tickets, and pieces of software that take us on this journey. We are not concerned whether the characters and their actions are real in themselves, only whether they are sufficiently convincing. We are buying experience – at second hand – and our principal concern is that it's quality experience. If we are really moved we will judge our money to be well spent.

The digital revolution has provided handy tools in getting the job done, certainly in the world of film and television where I work. The Titanic sank spectacularly, and the Gladiator performed to a roaring crowd of two hundred extras multiplied in cunning combinations again and again. Star Wars took us into experiences beyond the physical possibilities of any real theatre, and those great epics Exterminators I and II created staggering permutations of men and machines. We are not worried about whether it is real or not. It's credible, it's thrilling and it's fun. We don't care if Colin Radford points out that our engagement in these fictions involves us in "incoherence and inconsistency". Stargate, The Matrix, the Fifth Dimension, X-Men offer a roller-coaster ride through a world of fantasy, and we like it. Thank you SFX!

But work like this is not the end of the story. There will, of course, be more and more fantastic movies, but we must admit that they belong to a particular genre of entertainment. The creator does all the work, makes the story and the spectator watches, empathizes and is drawn along to the climax. This is traditional story-telling. The audience does not do anything other than get aroused.

Mutuality is reserved for that other great entertainment form of our age – the computer game. Here, narrative is interwoven with opportunities for the player to steer characters, choose routes and fight it out with monsters and villains. The power of this genre is well reflected by economics. The world computer game industry is worth more than Hollywood. It fascinates me to see how this world has attracted the opprobrium that was reserved for comics in my youth. Again and again I am told it is all sex and violence. People do not like to be told that these were also the selling points of the great classic epics, the Aeneid, Illiad and Odyssey, and that Titus Andronicus was Shakespeare's most successful play. Modern computer games – and of course I include all the console games Playstation, Nintendo and Dreamcast – play on a wide range of emotions. Riona's love for Squall in Final Fantasy 8 is a doomed love affair that brings in an avid audience of female teenagers who indulge serious emotions they are not getting from other media. Research reveals that many games players are affluent graduates in their twenties – and of both sexes.

What is it about computer games that is so appealing? Part of it, setting aside the imagination of their creators, is interactivity – the ability to be part of a story and yet have the illusion of influencing it. Another is the ability to move backwards and forwards in time, experiencing many smaller routines of engagement, thrills and satisfaction over the days and nights it is played. While most traditional forms of art demand nothing from the spectator apart from an entrance fee and a commitment to sit through it, computer games are about engagement. Television, faced with declining audiences and advertising revenue, is trying to learn its tricks (look at Banzai on Channel 4!). Across the water, if imitation is the sincerest form of flattery, Hollywood is now paying the tribute.

Engagement pays enormous dividends. The player who cares about, manipulates and champions a character "owns" him or her in the way children who impersonate their heroes and heroines own, and maybe, in their imagination, become them. They are like dolls or toys, as "tangible" as a software entity can be. This is the very soul of branding – concept burned deep into consciousness and memory.

The Interactivity which makes this possible is an inherent characteristic of digital technology with its freedom from linearity and ability to move speedily from point to point. Yet, the desire for this emancipation seems to me to be very old. In 1760, an Irish clergyman called Laurence Sterne, published the first two books of The Life and Opinions of Tristram Shandy, Gentleman. It is not so often read now, but in eighteenth century Britain it was a literary sensation, an uproarious, scandalous comedy, carefully designed and brilliantly executed. In its apparent disorder it can lay claim to being the first true work of non-linear story-telling in British fiction.

For those of you who don't know this extraordinary work, I can only point out certain features of its landscape. The first two hundred pages of this five hundred page novel concern the minutes leading up to the birth of the main character, events unrolling in a strange series of jerks and ellipses, punctuated by extensive and apparently irrelevant detours into fictional pieces of scholarship. The tone is entirely ironic. There are visual tricks – blanks, odd illustrations, and a black page when night falls. It is exactly like being involved in a superior computer game, but one which pays enormous benefits as when suddenly the odd, erratic and infuriating structure assumes significant form in the hilarious and muddy arrival of Dr Slop the man mid-wife.

Deep down the book is about obsession characterised as a hobby horse – a between-the legs metaphor for human fixation. Tristram Shandy conducts us on a wild dance through a series of obsessively explored ideas that collide, one moment opening to reveal satire, next closing and reforming in the shape of gigantic sexual metaphors. Sterne creates these obsessions as blocks in our own memory, and as he skips between them, we experience a sense of free movement. Even after we have put the book down we are being pulled backwards and forwards by the resonance of ideas.

Tristram Shandy emerges from a world in which the novel itself was a comparatively new form. Already, Sterne seems to want to blow the whole edifice of the novel form apart, challenging every notion of character, plot and authorial intention. Everything is conducted in a self-conscious manner, as if in a movie we were constantly being reminded of the true characters of the actors and shown the camera, lights and microphone as well as the inadequacies of the writer, and the philosophic hopelessness of the whole enterprise. I see the reflection from Laurence Sterne's hobby horse as a brilliant and anarchic light shining across the centuries and illuminating the possibilities of what we are doing today.

Deep in Sterne's technique is the knowledge that what is well lodged in the memory can be relied to surface when the relevant association calls it forth. This confidence in the power of memory was a feature of intellectual life that extended back into the earliest years of antiquity. In the latin text Ad Herrenium we have the story of how a nobleman called Scopas hired the poet Simonides of Ceos to chant a lyric poem in his honour at a feast. Simonides, did what he was asked, but also included a passage in praise of Castor and Pollux. Scopas, a recognisable figure even today (an accountant?) meanly told the poet that he would only pay him half what he was promised and that he would have to get the rest from the twin gods he praised. At that moment a messenger came in asking Simonides to step outside because two young men wanted to see him. Simonides rose from the banqueting table, and went outside. At that moment disaster struck. An earthquake shook the building and brought the roof of the banqueting hall crashing down. Masonry smashed onto the guests, crushing them beyond recognition. As relatives dragged the great stones away they did not know which body was which.

Simonides had been lucky, and now he showed that he also had a great gift. He was able to precisely remember the place at which each guest was sitting at table. Thus, he saved the day, and the relatives were able to identify their own. Simonides himself had, of course, been saved by the two twins Castor and Pollux who had called him outside and handsomely repaid his panegyric. This staggering escape clearly stayed in his mind, because he used his memory trick to invent the art of memory.

Frances Yates tells the story in The Art of Memory, which explores the history of the fascinating technique which was used for many centuries by orators to remember their speeches. It was an established part of the art of rhetoric and was actively used till comparatively recent times.

The trick was based on a mnemonic, which you may know. The practitioner learned to imprint the image of a places, or places in his memory. The location could be a building with many rooms, ornaments and statues. Images to recall the speech are "physically" set down in specific parts of the building. Then when the speech is given, the orator simply travels through the building, seeing the images and remembering what he has to say in the right order. This technique was adopted by many famous figures – Quintilian, Cicero, Metrodorus, with staggering effect. The amazing feature was that practitioners could travel backwards as well as forward. Augustine, an active pagan before becoming a saint was an adept but remembered the prodigious gift of a friend, Simplicius, who could recite the whole of Virgil both forwards and backwards.

The key to this extraordinary technique – and the interesting bit for us - is given by Aristotle. In his De anima he describes how the sensations of the five senses are worked upon by the imagination – literally the "thing that turns them into images" - and it is the images so formed which become the basis of intellectual thought. "The soul never thinks without a mental picture", he writes. It is the image making part of the soul that makes the work of the higher mental processes possible (had I known this forty years ago I would have had a stronger argument in favour of comic books). In the Ad Herrenium, the author advocates that teachers train their students in the art of creating images – the more striking and unusual they are the better they stick. How fundamental these ideas were to become we can judge by the fact that in the Phaedrus, Socrates ascribes them to the Egyptian god Thoth, later associated with Hermes, the central figure in the neo-platonic tradition who was to figure so strongly in the history of alchemy and science.

Whatever we feel about this technique today, there is no doubt that antiquity's use of image creation and manipulation for the purposes of remembering speeches and stories seems rather modern. Their coding uses icons rather than hexadecimals, but it enables a non-linear approach in which ideas can be recalled and assembled in any order at will. What is striking is the emphasis on using the imagination as a means of fixing ideas. The act of imagination literally becomes the trigger that creates memory. The subject does not passively receive ideas. He is active, using the creative imagination even when dealing with strictly factual entities.

This seems to me the essence of what makes interactivity important. By drawing on the viewer's imagination and choices, the interactive experience becomes his own. This use of the viewer's imagination has always been a fundamental building block in the hierarchy of artistic experience. Forms that are literal and show everything – simple movies, television, cartoons – rank less high than literature (which requires the reader's imagination to render characters and locations into specific images), poetry, and music (which works by the most abstract allusiveness).

Visual art can stand at the lowest point of literalness, or one of the highest. Sculpture has always been a form which has tempted complexity. In the Museum of St Peters in Rome there is a statue of an angel by Bernini which portrays the face of an astonishingly modern and pretty young woman. Walk 180 degrees and the soft lines harden into the ascetic face of a saint. When I first saw the statue I spent an hour walking backwards and forwards, watching the transformation and trying to understand how it was accomplished.

Rodin two centuries later tried a different approach. Obsessed with the study of movement, he sought to bring the work of Maret (that early motion capture photographer) into sculpture. In 1877 – 78, Rodin created his second great work, the

Walking Man, an armless, headless figure, which was later to become John the Baptist Preaching. By wrenching and distorting the legs, planting both feet flat on the ground, he created a number of profiles which give the illusion of movement. Walk around the figure, and one sees the weight shift onto one leg, and then back onto the other. A deliberate blow at the inertia of academic sculpture, this was not an abstract exercise in cleverness. Rodin was passionately involved in the human body as a vehicle for emotion. J. Cladel reports him saying: “The human body is a temple that marches. Like a temple, there is a central point around which the volumes are placed and expand. It is a moving architecture.” The Walking Man is a study of the simplest of actions turned into a human drama.

By comparison, Rodin’s controversial tribute to Balzac is complexity simplified to an image of significant ambiquity. Forty different studies testify to Rodin’s struggle to pay homage to France’s great writer. Given Rodin’s own nature it’s hardly surprising that he should finally discover the key in the great man’s sexuality. Ultimately, a pose of erotic excitement is hidden under a great cloak to produce what Al Elsen has called a “godlike visionary who belongs on a pedestal aloof from the crowd”. The life force is portrayed through physical excitement, and, as Steichen’s famous photograph revealed, the overall image resolves itself into a single, phallic silhouette.

In our own century we look at the complexity of Picasso’s line drawings and paintings, views of men and women from the back and side and top and bottom which all overlap and yet create a kind of coherence. There is a delight and revelling in the completeness of erotic detail. David Hockney looked at these works and said, rather drily: “Picasso wanted to see it all at once.”

Seeing it all at once is actually a characteristic of virtual reality. When we construct a virtual world it is there in full three-dimensionality, and the viewer has complete choice of viewpoint. We can walk round, through or over the objects. But this is quite trivial. One challenge is to motivate the viewer to exploring spaces which are complex and which make connections or create stories. But the real challenge is to construct something which is a space inhabited with ideas and characters and which has the richness to yield a range of narrative possibilities.

In 1910 a reclusive young Frenchman called Raymond Roussel published a novel called Impressions d’Afrique. Four years later he published another, Locus Solus. These two books were hailed as masterpieces by the surrealist movement which, astonishingly, Roussel claimed to have nothing to do with.

In Locus Solus, Canterel a wealthy scientist takes a party of guests on a tour of his extensive domain. They see a number of bizarre, and apparently inexplicable tableaux. One consists of a great crystal-faceted jar of liquid in which stands a beautiful woman clad in a skin-tight garment, hair flowing upwards and creating musical sounds with its movements. Various tiny emblematic figures perform small acts. A hairless cat swims through the liquid, puts his head into a pointed helmet and points it at a skinned face of Danton which hangs motionless. A touch of the helmet animates the face and Danton begins to speak. It is no wonder that Aragon called the author “president of the republic of dreams.”

Later Canterel explains the significance of all this in a rational way, and it is easy to see why he rejected the surrealist aesthetic. However, the power of Roussel's work lies in his almost photographic presentation of a three dimensional enigma, a tableau pregnant with the possibilities of individual interpretation. His detailed verbal descriptions are not always easy to take in, but they burn themselves into the imagination, and invite exploration.

Roussel used a deliberate technique in his writing. He would take a phrase containing two words each of which had a double meaning, and use the least likely meaning as the basis for this story. John Ashberry (in his introduction to Michel Foucault's book Death and the Labyrinth, The World of Raymond Roussel) tells how in Impressions d'Afrique he takes the phrase "maisons a espagnolettes" (house with window fasteners) and transmutes into a story about a a royal household or family descended from a pair of Spanish twin girls. He turned a line of Victor Hugo "Un vase tout rempli du vin de l'esperance" into "sept houx rampe lit Vesper" which he developed into a story of Handel using seven bunches of holly tied with different coloured ribbons to compose, on a bannister, the principal theme of his oratorio Vesper. Michel Leiris commented that Roussel is creating myths from a "disease" of language.

Whatever we think about compositional techniques like this, Roussel offers a fascinating insight into the possibilities of a new kind of fiction. A three-dimensional story space in which the reader can move freely across the surfaces from point to point, or directly across or through the centre. Each path brings a fresh association and relationship, and each is retraceable, repeatable or recombinable. This is literally story as architecture, in which the reader is as free to explore as in a building.

Now I think I can understand a complaint you might make. That I have taken you on a tour through some obscure and difficult territory. My excuse is that each of these little forays explores an aspect of what I see to be the possibilities of virtual story-telling today. I am not advocating the development of an obscure or elitist new art forms by examining these byways. It is in the very nature of new things that they are often are a little obscure. Picasso once showed an early cubist painting to a friend, and asked him what he thought. After pondering the question, and in some embarrassment – Picasso was, after all, a famous artist even then – the friend said that he had to say he found it rather ugly. Picasso did not explode, but beamed happily. "Of course it's ugly he said. Everything is ugly the first time you make it."

We all know what conventional story telling is. Characters, involved in situations of conflict, perform actions which constitute a plot. The forces which create the central crisis may come from within the main character, or from people or situations outside. The story teller – novelist, film director, choreographer – gives us various details of various kinds which build up a picture of what is happening. We become hooked on the situation, and stay with it wanting to know what finally happens. In no way do I deride this approach. There is something comforting about well-worn structures. Like sonata form, it has served well in the past, and will no doubt serve well again. But new tools have the habit of stimulating new ideas, and this conference is, I think, about trying to get excited about doing things in a new ways.

For me, perhaps the most exciting innovation over the next few years will be the development of virtual humans. There are already some impressive examples in computer games, but we have a long way to go before we can endow them with the illusion of credible human behaviour. In the film and television world, "character" and "personality" are key drivers, and we achieve this either by finding real people who are interesting, or actors who can impersonate the interesting behaviour of characters created by others. The idea that artificial humans could contribute anything worthwhile is incredible and indeed shocking to some people. Yet it is always in the regions of the unbelievable that we have to look for next big step forward.

There is of course a history of philosophic discussion about this subject. It starts with Descartes' assertion that we know other human beings have thoughts sensations, feelings and emotions because they have souls (therefore machines cannot) through Wittgenstein's discussion of the concept of pain (secondary) when a child plays with a doll, to Colin Radford's interesting questioning of the notion of sentience. Key to all this is the question whether a character has to be "real" – made of flesh and flood rather than plastic and wire, or even software – if we are to believe in it.

Clearly the answer here is no. There are millions of children throughout the world who have a relationship with Laura Croft, or Squall or Riona because their situations and characters are sufficiently credible to make them appear real. For people in the so called creative industries, these artificial humans are interesting because they are devices for story-telling with certain in-built advantages. Firstly they can be made to do things which normal humans would find difficult – leap, fly, fight and survive in impossible physical situations. Secondly, they are stylized, so they make a particular appeal to the imagination. Many of them "speak" using written dialogues like comics which are readily adaptable to different linguistic groups. Thirdly, they can be designed to satisfy an international aesthetic which will work as well in the Far and Middle East as in the USA or Europe. Fourthly, they can be controlled by the viewer. Fifthly, they do not cost as much as top rank film stars, and do not require subsidiary payments.

The physicality of these "avatars" is still relatively crude. Motion captured from the surface of bodies does not give a very accurate template to work from, while the behaviour of cloth on skin on muscle on bone is hard to render. Hair is another problem, and credible emotion difficult. Interestingly, all this does not seem to prevent games players enjoying existing characters enormously. But work is racing ahead in all these areas, and with a more complex set of interacting behaviour algorithms and more polygon rendering power, the day of the credible artificial human cannot be far away.

The range of behaviour and relationships in computers games is inevitably limited. Most games intersperse live action animated sequences with controllable virtual reality situations. The overall scenario is still highly planned, even with all its apparent options and freedoms. From a story telling point of view, the interesting question is whether we can ever move beyond this to a situation where the characters develop an autonomy in themselves.

For the last couple of years I have been involved in the design of a new software called Visions which enables comparatively unskilled operators to create complete prototypes of visual stories. I will not discuss it in detail because it is the subject of a separate paper, but essentially the creator designs and builds sets, fills it with relevant props and then "casts" artificial actors who can be made to move, speak and perform. All this has to be programmed by the creator, except for certain things. For example, in order to ease the way in which the artificial actors handle props, their hands are made to adopt a holding position relevant to a particular object, when the two are brought together. Bring a hand to a door handle and it will be held in one way. Bring it to a gun and it will do something else. Proximity triggers the inbuilt behaviour. To this extent, we have already achieved a limited autonomy beyond what we specifically ask the virtual actor to do.

The interesting long-term question is whether we will, in the future, be able to create an artificial human with characteristics of personality which can interact with the personality of another? If we can – and already we are quite advanced in the business of giving emotion to these artificial beings – we will be on the road to a form of story-telling in which dialogue, relationships and maybe even action will be generated organically, and beyond the imagination or control of the creator. Already we have software that can "learn" and build on interactions with operators. Is it impossible to imagine that at some stage our artificial humans will have the capacity to learn and grow as they interact with each other?

These will be challenging concepts for philosophers, and it would be nice to have Descartes, Wittgenstein and Radford back to hear what they have to say. For story-tellers, there will be fascinating issues about the design of prototypic situations that can "grow", as well as some interesting discussion about authorship rights and who gets paid for what. The science fiction possibilities are not to be discounted. After all, if the software is clever enough to have relationships, it might have the wits to find a way of getting out of its box.

Whether or not this "blue sky" vision is likely, or even possible, is not perhaps the purpose of this conference. Story-telling is a fundamental activity in our society, and finding new and more interesting stories is as important as developing the physical means to tell them. I have argued there have always been artists who have sought to "push the envelope". Laurence Sterne created comedy through unprecedented structural devices. Rodin could express the story of a man's struggle simply through a pair of distorted, anatomically impossible legs and an armless torso. Roussel built three-dimensional story tableaux. Meanwhile, at the simplest level, there have always been storytellers who have used the device of asking the listeners to suggest a new line for plot or character, and have resumed in that direction. Today, we can go on the internet and plug into a story like Online Caroline, and engage in a daily showerbath of narrative experiences with emails, SMS and even television.

It is however a fact, that technique does effect content. Today, the tools which are being developed for story-telling are of extraordinary sophistication yet increasingly accessible to unskilled operators. Technology has finally freed us from the constrictions of reality – we can go anywhere, physically, temporally, be big, small, inhabit any physical form however fantastic, be in any period of past or future. The question is whether these tools will feed back into the stories themselves. Colin Rushkov once said: “I’ve never enjoyed the virtual except as an idea.” Beyond the mechanics of story-telling, the impulse for narrative lies in the human mind. It is not only a question of what we can tell, or how we tell it, but what is worth telling. If our adventures in virtuality do nothing else, they may stimulate a re-examination of ancient questions, and cause us to look deeper into ourselves and how we interact with others and the world outside.

John Lee Hooker the famous exponent of those profoundly moving popular music stories stories, the Blues, died in the early summer of this year. He put it like this: “When Adam and Eve first saw each other, that’s when the blues started. No matter what anybody says, it all comes down to the same thing, a man and a woman, a broken heart, a broken home – you know what I mean”

Emotion, emotion, emotion. That’s what story-telling is about. And whether it uses real or virtual techniques, that is the heart of what we story-tellers have to deliver.

DocToon© – A Mediator in the Hospital of the XXIst Century

Benoît Labaye, Nadia Guérin, and Sébastien Dohogne

DocToon© Project
www.doctoon.com
info@doctoon.com

Abstract. The project DocToon© is an animation and rendering 3D system intended for paediatric wards to humanise the stay of hospitalised children and to help them to cope with the psychosocial problems they face. The principle of the technology is to animate, in real time, a little character, who can talk with the children in hospital, appearing to them via the television set in their room. An animator, in a room situated a little way from the children rooms, brings the character to life and speaks on his behalf, while seeing, via a two-way video the child with whom he is engaged in talking.

1 Introduction

Being in hospital is no laughing matter and can cause many anxieties for a child. On this basis, Paul Hannequart, doctor and managing director of a company active in the field of digital imaging and professional modelling, animation and rendering 3D, had the idea of creating DocToon©, a virtual character that communicates with children in hospital via their television set. The Regional Hospital Centre of the Citadelle (Liège) accommodated and supervised the pilot phase of DocToon©. The scientific part, carried out with the support of the Department of research, technology and energy of the Walloon Region, was undertaken by the University Hospital Centre in Sart-Tilman (Liège).

2 Technology

Unlike cartoons, DocToon© is a two-way interactive communication system. It is not only meant as being entertaining for the hospitalised children which it truly is but more importantly as integral part of the global care process. It is an integrated dialog forum that participates to the global well-being of the children by enabling communication on different topics whether related or not to the actual medical treatment the child undergoes.

The technical infrastructure is build around the control room where there two systems coexist. The first system, the Gabby© software, enables the character to be animated in real time and is used to bring DocToon© to live by making him move, speak, joke, talk story, … The second system, the Gestel© system, organises the communication between the Gabby© station and the television sets in the rooms.

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 171–180, 2001.

The system manages a network of Gestel© units that allow communication with children's rooms. The person who animates DocToon© dispatches the audio and video signals to the desired unit allowing communication between the selected unit and the control room. This management system is easy to control using user-friendly interfaces.

How does it work ? In the control room the animator of DocToon© sits in front of two PC monitors. When a child wants to talk to DocToon©, he calls him with the remote control and a new entry appears in the waiting list on a PC screen in the control room. The waiting list allows the animator to know the children who want to communicate with the character and how long they are waiting for him to visit them. In general the animator answers to the children in the chronological order of appearance although he can decide to select a particular child to talk to.

They are three manners to communicate with a child's room :

- A child calls and the animator answers : the communication is secure and only the persons present in this child's room can attend to or participate to the conversation. Nobody else is able to follow the dialog.
- The animator himself wants to broadcast a message to all the children rooms or to a subset : this is a one way communication as broadcast television.
- The animator would like to talk to a child who has not called him : at the first stage a "voice only" secure communication (without video relay for privacy issues) is initiated between character and the child. If the child wants to be seen by the character, the child must push on the "Call" button on the remote control.

Equipped with a headset, a microphone and a keyboard, the animator enters into conversation with the children and makes his character "come alive" and speak. To ensure a magic dimension to the system, the computer analyses the voice of the animator, transforms it in a "cartoon voice" in such a way the animator can not be recognised and applies it to the computer generated character in real time. The Gabby© software ensures that the character's lip movements are synchronised with the voice of the animator. Through a graphic tablet, the animator gives the character a full range of expressions (happy, laughing, surprised, crazy, hurry, ...) and movements that are pre-registered. The system is really easy to use and does not need technical skills. The animator just needs practice and feeling and can then be concentrated on his dialog with the child.

Each child has an infrared remote control with which he an call DocToon©, can choose the television channel and can control the sound volume. In each child's bedroom there is a Gestel© unit which is a small unit with camera and microphone that allows the animator to see and hear the child is in relation with. The Gestel© unit has also a display system to indicate the presence of the 3D character.

During the research phase, the communication between the animator's computer and the children's rooms took place via technology that used the hospital's television distribution cables. This technology, although remarkably efficient, nevertheless presented certain disadvantages (risks of interference with other channels distributing other programmes via the same cable, no possibility of installing the system in hospitals without television distribution, the possible requirement to adapt each installation according to the frequencies already in use).

For all these reasons, it was decided to start working on a different technological approach, based on an autonomous cable, independent of any other installation within the hospital. The company's electronics experts also developed a completely new design for the Gestel© communication unit, which was more compact, more efficient and which fully respected EC standards. A trial installation successfully passed all the tests laid down by the specifications, enabling in particular communication at a distance up to 300 metres without loss of quality. In order to ensure a higher distance, a repeater will be created to receive, to amplify and then to send back the audio and video signal to longer distance cable.

The technology allows further developments to be envisaged : accompanying the children into the operating theatre, announcing meal times, inaugurating story time, providing monitoring of the child at home …

3 Virtual Storytelling at Hospital

3.1 Restoring a Smile Is Already a Step towards Recovery

Just imagine a child in a hospital room. Using a remote control he can access all the different channels on his own personal television. One of these channels is something rather special. When the child selects it a little cartoon character appears on the television screen and asks the child what he wants. The child can then start talking to the amusing little three dimension character.

What do they talk about? Well that's a secret, of course…. but they talk about doctors, hospital, friends, school, family, etc… They have a good laugh, share a joke… That's what DocToon© is. Someone to confide in when the child wants to. Someone who is there and who never comes and 'does something' to the hospitalised child, but just listens, talks, plays, has a laugh. DocToon© won't theorise about the meaning of life or the illness itself. He is not there to replace the doctor or nurse. His role is more as a facilitator to encourage communication between the child and the adults looking after him. By talking to the child, he helps him to face up to and cope with his anxieties and the "existential" problems he encounters during his stay in hospital. He's a funny, cheeky, understanding, reassuring interactive friend. DocToon© is a concrete example of virtual storytelling imagined to solve the communication problem in an original way.

3.2 Ethics and Privacy

In order to respect ethics and privacy the ethical charter covers all persons (doctors, psychologists, nurses, social workers…) who work with DocToon© and stipulates amongst others the personal and technical competence of the animator, the limits of the animator's competence (neither doctor, nor nurse). The charter also stipulates that the DocToon© system and its objective will be the subject of a systematic presentation to any child, that only one animator is authorised to communicate with a given child at different times (the animator for any one child may not be changed), ...

4 Experiment[1]

4.1 Aim, Methods, and Observations

The experiment use of DocToon© in Paediatric Ward of the Citadel Regional Hospital Centre, Liège, began on April 1998. The aim of the experiment was to assess the benefits and risks of adding a virtual character (DocToon©) to a paediatric nursing and medical team.

The observers were: children, their parents, medical staff of ward 57 (doctors, nurses, social workers), child psychologist animator of DocToon© and two external observers.

The observations were collected as follows:
- Observation of children : an adult speaks with them of their stay at hospital.
- Observations of parent and medical team of ward 57: through anonymous open-ended questionnaires.
- Observations of child psychologist animator of DocToon© : through a discussion with one external observer.
- Observations of the two external observers : writing up of the observations collected and of their own impressions.

The content and psychosocial, social and emotional environment of the conversation between DocToon© and the children seemed to open up perspectives that we had not even imagined.

The observations were:
- DocToon© elicited curiosity, interest and enthusiasm in all the hospitalised children. The children clamoured for him.
- The children became strongly attached to DocToon©.
- Corollary of the first two points: Adults, whether the psychologist or someone else, were accepted more readily by the children when their actions or presence was mediated by DocToon©.
- DocToon© amused the children and created something magic in his relationship with them.
- The relationship with DocToon© reduced the children's stress.
- The relationship with DocToon© changed the child's perception of her/his situation.
- The relationship with DocToon© changed the family dynamics around the hospitalised child.
- DocToon© changed the mood in the nursing and medical team; he brought them joy and good humour.
- The relationship with DocToon© facilitated some usually difficult medical acts. It seemed as though the relationship with DocToon© reduced the pain that was felt.
- Generally speaking, the relationship with DocToon© increased patient compliance.

[1] By Lambert Maréchal, PHD

- In some situations the relationship with DocToon© reduced the hospitalised child's feeling of loneliness.
- DocToon© was sometimes an escape hatch for the child's aggressiveness and violence.
- No drawback was seen, other than criticisms related to technical flaws (sound quality, remote controller's reliability, etc.), and DocToon©'s visiting hours (not frequent enough).

4.2 Hypotheses about the Mechanisms That Might Explain the Virtual Character's Efficacy and More Specifically the Advantages That We Observed

4.2.1 First Hypothesis

As Piaget[2] showed in his remarkable studies, children's representations of the world are initially devoid of distinctions. At the start of their lives, children are 'realistic'. They assume that thoughts are linked to their objects, that names are linked to the things named, and that dreams are external. Their realism consists of a spontaneous, immediate tendency to confuse the sign and the signified, internal and external, and psychic and physical. The consequences of this realism are twofold. First, the boundary between the ego and outside world is much fuzzier for children than for adults. Second, this realism is continued through spontaneous 'participations' and magical attitudes.

Self-awareness is thought to result from a dissociation from reality. The child achieves this dissociation by differentiating the others' points of view from her/his own point of view. In the beginning, the child considers all representations to be absolute, as having the spirit enter the thing itself. Only gradually do children conceive of representations as being relative to a given point of view. The child starts by confusing her/his ego and the world (the subject's point of view and the outside given), then differentiates her/his own point of view from the other possible points of view. To the extent that s/he is unaware of the subjectivity of her/his point of view, s/he thinks s/he is at the centre of the universe. This gives rise to a set of quasi-magical, animist, finalistic conceptions of the world: the child believes that the sun, moon and clouds follow her/him, that things are always as s/he sees them. So, up to the age of about 11, thinking is equivalent to speaking, and speaking consists in acting on the things themselves through words. Words participate in a way in the named things, as do the voices that utter them. There is realism and realism due to a perpetual confusion between the subject and object, between the inner and outer world.

The animism of children is not theoretical, i.e., intended to explain phenomena. It is emotional ('the stars are interested in us'). Up to the age of 7 or 8 children refuse to accept that things do what they want to because they believe that everything's will is governed by a moral law built around the principle that everything is done for the good of people. The first notion of physical determinism wells up around the age of 7 or 8. This new idea is slow to be systematised and not until the age of 11 or 12 will it replace the idea of a moral rule in the child's physics once and for all.

---

[2] Piaget J., La représentation du monde chez l'enfant. PUF 8eme éd. 1966.

The psychoanalytic approach stressed the importance of the imaginary in children's lives. The primary process consists in creating a mental picture of the desired object. This mental picture is a hallucinatory satisfaction of a need. If the game and imitative representation suffice, at least in the early years, these approach believes that this is the consequence of the exaggerated value that the child attaches to her/his desire. Relieving the tension through the primary process offers temporary satisfaction of the need. Thereafter, the drive will continue to well up, whereas the hallucination will not be efficacious forever. At that point the child will search for a real object to provide satisfaction. At that point, reality, that is to say, the outside world will have to be taken into consideration.

It is important to stress that children almost never speak about their visions and representations of the world. At first, this silence is predicated by the uselessness of such speech. Since the child assumes that everyone thinks like s/he does, why should s/he explain to others what they already know? Later, when the child begins to realise that other people - her/his parents - do not necessarily think like s/he does, s/he prefers to remain silent, for having omnipotent beings challenge one's thoughts and personal convictions is always upsetting and a source of insecurity.

4.2.2 Second Hypothesis

For the child, growing up corresponds partially to a narcissistic wound. The narcissistic wound with which the 'ill' child must cope in the course of growing up is even more difficult to bear than that of a healthy child. By affecting the child's physical and, to a variable extent, intellectual abilities, disease and resulting hospitalisations hobble the child's development. They disrupt identity-building and shake her/his self-confidence. The more or less major break with the usual social environment - up to and including isolation - reinforces this disruption.

Illness deeply affects the self-image that the child is constructing. So, for example, when Jeremy - a child hospitalised in Paediatric Ward 57 - was asked by DocToon© to introduce himself, his first sentence was, 'I was born and I had diabetes.' The sick child must cope with two narcissistic wounds. Like all children, s/he must give up the fantasised omnipotence of her/his ideas and desires. On top of this, s/he must also renounce the happiness that, in her/his fantasies, is linked to the 'good health' of which s/he is deprived. In this particular circumstance the 'sick' child needs reassurance more than ever. S/he has an intense need to understand and give meaning to her/his life.

4.2.3 Third Hypothesis

The first observations that we have made are still too fragmentary. However, they suggest that there might be three age-related stages in the way that children interpret the degree of 'reality' of their dialogues with the virtual character.

In the first stage, the child is not at all amazed by the fact that DocToon© sees and speaks to her/him. Indeed, it is normal for DocToon© to see and know the same things as the child, since all representations at that stage are absolute and the child has not yet grasped the relativity of points of view.

In the second stage, the child is very excited about DocToon©'s appearing on the screen and seeing and talking to her/him. 'It isn't normal,' and, to use one child's words, 'There's a trick there.' The child looks around and, when s/he put her/his finger on the Gestel's lens and thus prevents DocToon©'s animator from seeing her/him, s/he is happy, reassured, and perhaps has fortified her/his new understanding of the world.

The third stage might be one of rejecting the world of childhood, which might also include nostalgia for a 'paradise lost'. It is as if the child begins to fear and regret the fact that wonders might not exist. 'DocToon© isn't real, but you mustn't say that'. As Geoffrey, who is just over 8 years old explained it to us, 'DocToon© is a man wearing a disguise. That's obvious. Because Santa Klaus, Father Christmas, and the (flying) bells don't exist; they are disguises. Yes, but that will make kids unhappy. If DocToon© tells the kids the truth, it will be very hard on the kids, because it will make them very sad. They love DocToon© so much!'

4.2.4 Fourth Hypothesis

One of the reason's for children's fascination with DocToon© might be that he enables the imaginary to be accepted again in the child's world. The virtual character is given to the children by the same institution that gives them the care that is considered effective in the adults' world.

4.3 Approach for Specific Situations

Certain pathological configurations and certain situations must by all accounts be handled with caution. Thus, despite the reservations that could be raised at first sight, it is difficult to assess the real impact, positive or negative, that the use of DocToon© might have in child psychiatry, for example, without detailed scientific research. In particular, it would be useful to study whether the virtual character of DocToon© is likely to disrupt in a negative manner the imagination of psychotic or seriously neurotic children. Are there certain circumstances in which, on the contrary, this virtual character could be used for just the right therapeutic approach? This delicate research would require a competent team, familiar with the population concerned, and a detailed analysis of the actual conditions in which it might be undertaken.

Other specific situations merit research without prejudice to determine the extent to which the intervention of DocToon© could be of benefit. We are thinking, for example, of children isolated in sterile rooms and deprived of the essential sensory contacts.

We are also thinking of company for children living in seriously disturbed family situations (domestic violence, incest, divorce conflicts...). In these situations, there is clearly no question of attempting to or considering substituting DocToon© for *face to face* human contact, but rather to study whether the parallel use of DocToon© can enable verbal or non-verbal expression of things that might not otherwise be expressed.

4.4 Further Research and Investigation

Regardless of any other consideration, the success of DocToon© can be assessed empirically through the enthusiasm of the children, the frequency of their calls to

DocToon©, the obvious pleasure they have when in contact with him, the positive comments from the doctors and nurses, the evidence from parents and the appeal for all external visitors. The promoters of the experiment are aware that these empirical impressions cannot suffice, and that, in a field as sensitive and of such high human value as that of a sick child, a serious scientific assessment should be able to be undertaken, which would analyse the operating mechanisms of DocToon©, its contributions in terms of well-being and assistance with treatment and its possible limitations. They are also conscious that this research could only be convincing on the two-fold condition that sufficient resources were available and that it was totally independent of the designers.

5 Summary and Conclusions

5.1 Who Are You ?

The novelty of communication between a TV set and individual does not seem to pose any positioning problems for children, at least for under-12s. In our case, the children did not seem to be upset or even truly astonished by seeing a cartoon-like character talk to them through a TV set. They are obviously aware of the fact that this character does not exist in flesh and blood. DocToon© is not a material, tangible being. On the contrary, they enter into the magical dimension of the undertaking with amazing ease. Their imagination immediately plot out the world in which DocToon© will exist and take his place. They do not wonder about the technical and material components of the communication and do not ask for explanations of how it all works. The 'tricks' don't interest them. They immediately enter the 'game', in the noblest sense of the word, *i.e.*, a world in which the imagination creates entities and connects them through dynamic relationships in which the marvellous and pleasure act directly on the sensations of 'being in the world' (on suffering, for example, tension, stress, and so on).

The first step in this 'work' is usually, and rather naturally, to set the imaginary boundaries of the game by establishing, through a series of questions and answers, a sort of business card for DocToon©: Where do you come from? How old are you? Where do you live? etc. The 'puppeteer's' answers leave sufficient room for the 'marvellous' while giving the child enough elements to build a topology in which s/he can move about.

The question of DocToon©'s age is extremely important. Like all cartoon characters, like Peter Pan and Snow White, DocToon© doesn't really have an age. His size, appearance, and language mirror a series of childhood characteristics that makes it possible for the child to project an age close to her/his own on DocToon©. Still, Peter Pan and Snow White are also more than children. Their extraordinary experiences, courage, ingenuity, cleverness, and ability to cope with the world and its dangers give the child an idealised image of her/himself into which s/he can project her/himself. This image helps the child to face her/his own tensions and reinforces the child's self-image, as documented in the work of Bruno Bettelheim.[3] In

[3] Bettelheim, B. The Psychoanalysis of Fairy Tales.

DocToon©'s case, something similar occurs. He knows everything and everyone in the hospital, knows and can talk about a lot of things.

Like Tintin, for example, he also does not have any visible guardians. There aren't any 'parents' on his flying saucer and planet. When he mentions his 'Dad', it is more in the role of a creator than an authority. DocToon© is thus a sort of free child. He is clever and intelligent, can speak with both adults and children, and determines himself. He is an idealised child who, through imagination and projections, can doubtless reassure and bolster the child's self-confidence.

Next, two essential features not shared with traditional cartoon characters are vital to the richness of DocToon©'s persona. The first one is clearly that fact that *'he talks to me!'* He is not a silent puppet, a hero locked into a story that goes round and round in an endless loop, or an adult wearing a disguise. He is a true Martian from animation who 'sees me and talks to me!' There is a totally new, fabulous interaction with DocToon©, somewhat as if the child could call Peter Pan over to talk things over. The second, subtler, feature is the fact that, since he enters the child's reality, he also has to grapple with limitations, impossibilities, misunderstanding, and ignorance of things in an unfamiliar universe. DocToon© cannot leave the computer, television set, and network. He can't walk in the street, go to school, or play soccer, and there are lots more things that are impossible for him. So, in contrast to the perfection of a cartoon hero who is closed into himself, since he cannot escape his story, DocToon©'s interactivity makes him an open, dynamic person, in a way, someone who is fragile and imperfect. The child can also teach him things, inform him, correct him when he makes mistakes. The relationship is not one-way. The child is also an initiator or protector of DocToon© in the unfolding friendship.

5.2 Real and Imaginary

These two characteristics doubtless explain why DocToon© creates a very special relationship between the real and imaginary worlds. Whereas traditional animated films mobilise primarily the imagination, which can put the child physically in a state close to apathy and mentally in a state close to hypnosis, things are fundamentally different with DocToon©. In dealing with DocToon© the child is active. Moreover, when her/his condition allows it, s/he tends to move about, jump up, turn over, look for accessories in her/his room, or involve bystanders in the conversation. These are all signs of intense inner activity. We even think that DocToon©'s oft-observed efficacy in relieving pain or stress is less a matter of anaesthesia (the *morphine* of words) than the result of mobilisation of the child's imagination, emotions, and senses in such a captivating activity that everything else is relegated to the backseat of the child's consciousness.

This relationship between the real and imagined is quickly materialised in the substance of the conversations. The talk with DocToon© typically concerns the child's experiences and daily routines. DocToon© gets the child to talk about her/his family, parents, siblings, school, pals, games, tastes, and obviously her/his illness or the reasons for her/his hospitalisation. Bringing these daily matters into a world of fun, into a relationship that mobilises mental and imaginative activity, triggers a shift in the child's view of her/his daily life. DocToon© reflects only the positive or reassuring emotions and assessments of who the child is and where s/he lives. However, this *mirroring* is not the work of an adult in a position of authority, an educator, even someone with a potentially threatening status, such as a doctor, nurse, even the real-life psychologist. The person in whom the ego projects itself and looks at itself is a sort of magical pal with whom life (for we are talking about everyday real life) becomes a game. DocToon© never makes judgements or demands, never utters deprecatory remarks or recrimination. Through DocToon©'s eyes, everything that the child experiences, including disease, hospitalisation, and the burden of anxiety or guilt that may be connected thereto, becomes easier to objectivise, makes sense. Whereas relations with other adults, including and perhaps first and foremost the child's parents, are characterised by the interaction of facts linked to obedience, submission, the fear disappointing or being abandoned, and complex emotional ties, with DocToon© the child feels both unhampered and appreciated, without obligations or conditions.

5.3 He's a Friend …

All of this may explain why a true friendship develops between DocToon© and the children. This relationship was revealed by various elements, for example, the children's frequent desire to give DocToon© drawings, to give him presents, to leave him boxes of chocolates or other goodies upon being discharged, even asking DocToon© to write to them when they've left the hospital! This relationship is truly intense, and give the hospital stay a very special colouring. The prospect of hospitalisation is accompanied rather naturally by negative, upsetting emotions linked to anxiety, real or projected suffering, and the fear of the unknown. DocToon©'s presence throughout the hospital stay changes things fundamentally. His faithful, unconditional presence causes something kindly reassuring, funny, and appreciative to happen. This adds a heretofore unknown human dimension to the hospital stay. DocToon© mobilises the child's imagination in a positive way around everything that can reinforce the child's narcissism, which takes a beating from the illness and suffering. So, through a relationship that is totally devoid of all negative tension, DocToon© may in this way objectivise, complement, and dedramatise the attention and affection focused on the child by her/his parents and/or nursing staff. DocToon©, a human-computer interface, illustrates how the Virtual Reality can offer new tools to capture and interacts on the imaginary environment.

The Interplay between *Form*, *Story,* and *History:* The Use of Narrative in Cultural and Educational Virtual Reality

Maria Roussou

Foundation of the Hellenic World
"254 Pireos St., Tavros 17778
Athens, Greece
mroussou@hol.gr

Abstract. This paper attempts to review examples of the use of storytelling and narrative in immersive virtual reality worlds. Particular attention is given to the way narrative is incorporated in artistic, cultural, and educational applications through the development of specific sensory and perceptual experiences that are based on characteristics inherent to virtual reality, such as immersion, interactivity, representation, and illusion. Narrative development is considered on three axes: form (visual representation), story (emotional involvement), and history (authenticated cultural content) and how these can come together.

1 Introduction

Storytelling is a familiar and ubiquitous part of everyday life; a form of communication and shared experience since antiquity. Stories serve social, cognitive, emotional and expressive functions. In the social and pedagogical sense, storytelling can serve as an effective representation for learning, whether it involves experiencing a pre-determined narrative script or constructing ones own story. In terms of its cognitive function, the structure and dramatic tension of a narrative creates expectation which is satisfied upon resolution of the story, and aids in planning, reconstructing, illustrating, and summarising abstract concepts. Most of all, it is the affective function of narrative, as explored through artistic genres such as literature, theatre and cinema, and plot-based media such as games, that fascinates immerses, and lends its form to interactive media and virtual reality productions.

Characters and story structures are not completely unexplored concepts in interactive media, however they are undermined by the dominance and almost exclusive development of the visual form. Interactive virtual reality applications, in particular, have mostly focused on the construction of objects and spaces, but not stories that tie them together. This is partly due to the technical limitations of the tools. However, another reason why stories are not incorporated in the design of virtual environments can be explained by the fact that the emerging field of virtual reality is still uncharted territory and its actual use as an artistic, educational, and cultural medium is largely overlooked or unexplored.

In this paper, we attempt to look at examples of storytelling and narrative in immersive virtual reality worlds and analyse the form of narrative as it pertains to art,

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 181–190, 2001.

culture, and education through the development of specific sensory and perceptual experiences that incorporate immersion, interactivity, representation, and illusion.

2 Interactivity and Narrative in Virtual Reality

Interactivity is a raison d'être of a virtual reality world. Ryan claims that our culture, formerly one of immersive ideals, is now a culture more concerned with interactivity [15]. Rokeby writes that interactivity's promise is that the experience can be something you do rather than something you are given [11]. Today's virtual reality interfaces, due to their immersive and interactive qualities, are designed in such a way that the user is literally placed in the scene and is actively engaged with the surrounding environment. The development of systems such as the CAVE™ presents one of the better examples in this direction.

However, most people do not know or understand how to deal with interactive computer-based environments, let alone with interactivity in immersive and, in many cases, complex virtual worlds. The virtual experience can be disorienting, unnatural, and difficult to become part of, even if the technology used is as simple and natural as current development allows. Numerous observations of children, adults, single viewers, groups, novice and even expert users in virtual experiences have indicated that interactivity is not necessarily all that matters. Rather, realistic simulation and a fascinating story to complement it seem to make up the formula of illusion.

In any case, it is evident that most artistic VR applications strive to achieve a high degree of interactivity. Despite this fact, the first artistic explorations of virtual reality produced abstract worlds of non-associated objects or spaces. On a parallel but antithetical level, cultural VR experiences have become synonymous to passive walkthroughs of realistic (technology permitting) yet simplified recreations of architectural worlds. In both instances, the dominance of form and the lack of structure indicate that little or no conception of narrative existed in the design and development of these virtual environments.

It is a fact that, while many narrative-based interactive art installations, two-dimensional computer-supported storytelling environments for children [16] or computer role-playing games [9] have been developed, narrative has just begun appearing extensively as a theme in interactive virtual environments. So far, virtual worlds that are based on narrative theory and structure are almost exclusively developed by artists or linked to the area of Interactive Fiction. These systems explore narrative more in the sense of space and time [7], and less in the direction of plot and character development as encountered in traditional literary narratives or emotional interactive drama, thus ignoring the addition of the user that interactivity places as the core of the equation. Laurel proposed the use of drama as a metaphor for computer interface design by placing the user in the role of both spectator and director [7]. However, placing the user in an active role complicates the conventional narrative patterns of author/storyteller to reader/listener/receiver.

Let us look at some examples of immersive virtual reality works that have more or less been engaged in forms of narrative and storytelling from their design phases to the final outcome. Although their approaches to narrative and form vary greatly, Benayoun's *World Skin* [3] and Fischnaller's *Multi Mega Book* [5] are both CAVE®

virtual reality experiences in which the synergy of form, story, and underlying concept serves immersion and illusion to the highest degree.

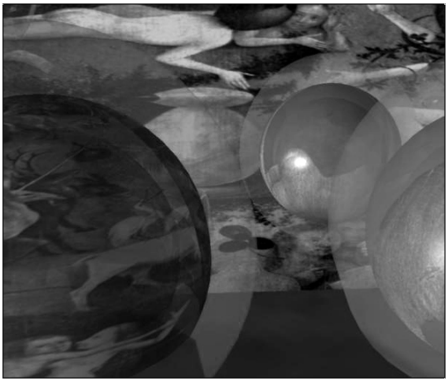


Fig. 1. Mitologies: cinematic narrative in VR, where the user is allowed choices that determine the path taken in a virtual labyrinth. Courtesy of M. Roussos, H. Bizri, 1997. http:// www.evl.uic.edu/mariar/MFA/MITOLOGIES/

Mitologies, a virtual reality artwork created for a CAVE-based environment, is an attempt to employ traditional narrative content and structure to a virtual experience (Fig. 1). The film-like mise-en-scene used was selected both for its familiarity with the viewers and as a mode of expression. The narrative draws inspiration from a pool of mythological and medieval literary and artistic sources, taking a different approach to virtual narrative structure by almost ignoring interactivity. The thematic content of Mitologies is loosely based on the Cretan myth of the Minotaur, the Apocalypse, or Revelations, of St. John, Dante's Inferno, Durer's woodcuts after the Apocalypse, and Borges' Library of Babel. Music from Wagner's Der Ring Des Nibelungen is used as a motif to structure the narrative. The work explores the enigmatic relationships between these sources and captures them in a mise-en-scene that is rooted in the illusionistic narrative tradition of other media, such as cinema. Although created and exhibited in a virtual reality platform that allowed for a high degree of interactivity (a CAVE®), in most cases the audience of Mitologies has no control. The cinematic narrative form preserves itself through the continuous slow pace and progression achieved from one scene to the other. The virtual journey through a labyrinth presents its visitors with a narrow range of choices, yet all choices are in essence illusory, as they ultimately lead to the same final confrontation with the minotaur, the fall through a trap door, and the return back onto the boat from which the experience begun, thus completing a circular journey [12].

On the other hand, *The Thing* [1] engages the user in interactivity through constant "conversation" with a virtual character rich in changing emotional states. The work is structured in three acts in order to take advantage of narrative tools like pacing, surprise and movement through time. In order for the story to progress, the user must engage in activities and respond to the character's requests by dancing, moving,

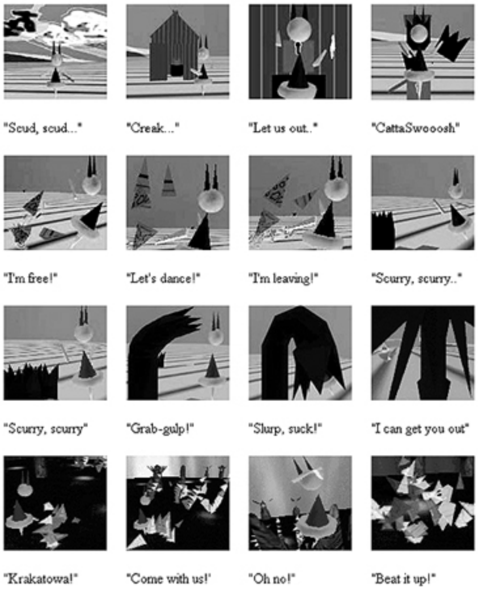


Fig. 2. An "activity" storyboard of *The Thing Growing*. Courtesy of Josephine Anstey, 2000.

selecting objects, or performing actions (Fig. 2). *The Thing* provides us with an example where interactivity and narrative are closely intertwined: storytelling serves as a driving force for a highly interactive experience and, vice versa, interaction between real and virtual character, plot, and emotion becomes central to the formation of the story.

The above two approaches to narrative in virtual reality are situated at the two different ends of the interactivity - immersion spectrum. *Mitologies* employs high quality visually complex scenes that take advantage of the immersive qualities of the medium to the expense of interactivity. The cinematic form of narrative is familiar and safe. It does not allow much exploration of the narrative form and does not require much activity on the part of the user (thus also eliminating the need to train the user). *The Thing* bases all of its power on the interactive by maintaining a simple visual and aesthetic form. The visuals are used to set the scene rather than define the artistic process, while the constant demand for interaction between the participant and the virtual presence (character) help to almost ignore the surroundings. Despite this fact, the participant's discourse with the "Thing" becomes so involved that a strong sense of immersion is also achieved.

3 Virtual Storytelling in the Formation of Cultural Experiences

Virtual reality technology both in theory and in practice is increasingly being considered and supported for the new possibilities it can offer to cultural heritage representation and education. Museums, as the main representative authorities of cultural and historical content, are adapting more and more interactive hands-on techniques and virtual technologies for use in exhibitions and public programs.

As museums become more open and involved with interactive technologies, their conception of the audience as active participant, or maybe even creator of the work emerges and the creation of "experiences" and themed exhibitions comes to the fore. Fantasy and illusion are key elements to the construction of experience, as is a story told well. In this sense, storytelling is included in different types of cultural virtual worlds to serve the idea of an active experience in the sense of an "expanded metacinema", to borrow Peter Greenaway's words. Whilst referring to cinema, film director Greenaway suggests to integrate all manner of sophisticated cultural languages into a three-dimensional form with "stimulus for all five senses where the viewer is not passively seated, can create his or her own time-frame of attention and can (as good as) touch the objects he is viewing and certainly have a more physical / virtual relationship with them" (cited in [10]).

While museum audiences do not expect today's museums to have reached the level of sensorial richness illustrated by Greenaway's vision, they do expect the museum to tell stories. Museums tell stories through the collection, informed selection and meaningful display of artefacts and the use of explanatory visual and narrative motifs in their exhibits and in the spaces between exhibits. The stories expected and inferred through the exhibits are part of an interpretative process that provides cohesion for the exhibited content. This interpretative process is at the core of the museum as an unassailable institutional authority and remains the most significant factor that differentiates museums as informal education spaces from public recreational venues, such as theme parks. In other words, authenticity is both an effect that exhibit makers strive to achieve and an experience that audiences come to expect from museums. It is thus crucial for museums to preserve this context of knowledge and credibility while providing memorable experiences that can tell the stories and, ideally, suspend disbelief. Suspending disbelief is one of the key aspects of narrative engagement and perhaps the most central goal of an immersive virtual environment. But how does authenticated cultural content that traditionally involves a research-based process with multiple perspectives relate to the emotional and dramatic patterns of narrative?

At the Foundation of the Hellenic World (FHW), a cultural heritage institution located in Athens Greece, virtual reality is used both as an educational/recreational tool and as an instrument of historic research, simulation, and reconstruction. The FHW develops its own cultural and educational virtual reality programs that are shown to the public in the cultural center's two immersive VR exhibits/theaters: the "Magic Screen" (an ImmersaDesk™) and the "Kivotos" (a cubic immersive display for up to 10 people). The programs range from highly detailed reconstructions of ancient cities that can be experienced as they were in antiquity to interactive educational programs that require active visitor participation [6].

All productions have an embedded sense of narrative. In some of these programs this is presented in a more literal and obvious form through narration, while in others, storytelling is implied through the interaction with the virtual environment and the completion of tasks with a concrete goal. In "the Magical Wardrobe" program, for instance, young users can select a garment from a set of virtual costumes, each from a different period of Hellenic history, and by "wearing" it, be transported to the corresponding time period of the past. Once in this fairytale land of colorful scenery and virtual characters, the task is to search for costume elements and accessories in order to help the virtual people of the specific time period prepare to take part in a celebration. The process of searching for, discovering and identifying different costumes provokes inquiry that can lead to knowledge of the cultural, sociological, and political importance of costume at the time (Fig. 3).

Fig. 3. Interactive storytelling based on historical content: a museum vr experience where children actively participate in exploration through a story with a concrete educational goal. Courtesy of the Foundation of the Hellenic World ©2000. http://www.fhw.gr/

The use of virtual representation for cultural and historical content, as implemented in this context, has proved to be a strong public attraction and engagement force that can redefine the relationships between the audience, the venue, the virtual representation and the real object or historical fact. However, the difficulties of representation entailed by showing and telling stories about sequences of historical events over time or space are immediately apparent and no less daunting than those entailed in the task of developing virtual narratives. The constant struggle to merge historical accuracy, aesthetic pleasure, and engaging educational value has been an even greater challenge than the technical difficulties in achieving high performance and quality real-time graphics. In this case, the synthesis of form (representation), story (narrative), and history (cultural content) by the museum (authority) is a difficult and sensitive fusion to achieve. Especially as its ultimate aim should be not to teach in the didactic sense, but to encourage the exhibit's visitors to question what they experience and to engage in "contradiction, confusion, and multiplicity of representations" inherent in the display of historical and museum content [10] while at the same time avoiding the danger of collapsing time periods into an attempt to redefine them as part of a confusing and fragmented experience.

In this sense, successful approaches to virtual storytelling from the cultural perspective should exemplify this interplay of concepts such as historical accuracy, educational efficacy, high motivational & engagement level, quality visitor experience, and seemless, natural, and customised modes of interaction.

A way to achieve this cultural narrative context is to complement the virtual experience with the undeniable power of human storytellers. The role of the museum educator, guide or facilitator has been critical in helping the audience build bridges between these different perspectives to gain a deeper understanding of the content. The use of museum educators, archeologists, and teachers as museum guides in the virtual experience not only adds value to the story but can promise the development of unique stories every time. The museum thus maintains the potential for multiple different experiences that respond to visitor needs rather than a single, repeatedly identical experience. Different people employ different processes and have different comfort-levels with the technology. The multiplicity of approaches also means that the visitor experience depends on the skills of the storyteller/guide in the sense that even unpredictable external reasons ("having a bad day") can dramatically determine the quality of the experience making it inconsistent. These processes are reflected in the formation of the visitor experience and the methods of structuring the interactive experience to encourage new forms of interactivity under the context of a story. The stories vary as guide preferences and capabilities vary: some may choose to keep exclusive control of the interface and others to share the controller between all the visitors. Some may prefer to direct the experience, others to suggest courses of action to the visitors. Some encourage interaction while others prefer a more structured experience. Some use the experience as a way of generating questions from the visitors others as a vehicle to dramatic improvisation and magic.

The use of intelligent agents for storytelling purposes in virtual environments presents an attempt to simulate these human qualities.

4 Avatars and Intelligent Agents as Educational Aids or Storytelling Characters

If characters are critical to story plot, then the development of avatars and intelligent agents as simulations of life-like characters is at the crux of development for narrative and storytelling functions in virtual worlds. Characters in virtual worlds draw on codes heavily used and tested by the masters of illusory entertainment experiences. Their role is one of delivering anthropomorphism, embodiment, and believability to a virtual experience.

Incorporating story and characters requires the development of more "intelligent" computational models in virtual reality systems. Recent advances in the field of artificial intelligence include the development of agents, artificial creatures incorporating a set of human-like behaviors, as well as the exploration of plot and story structures, which may emerge from the interaction between these agents [2]. Despite the interesting technical developments in the direction of natural language processing, speech generation and synthesis, gesture, lip syncing, facial expressions, etc, these programmed agents have far to go before they can successfully simulate a perceptual, cognitive, or emotional level that may produce consistent and coherent narrative. Technical limitations have not allowed as yet for the development of agents that are intelligent enough to respond to the human users' wealth of emotional states and improvisational behavior, in order to construct a meaningful interactive story. Despite the attempts to create believable software-based intelligent agents, from the first ELIZA to the present computer game characters, agents in narrative-based virtual environments can convince only for short, fragmented and simple, or relatively plotless stories [8]. A common use of agents in cultural heritage applications takes the form of virtual guides that carry a predetermined set of actions and prerecorded speech. More inventive variations explore the aesthetic options by replacing realism with more abstract form, speech with gesture, etc. (Fig. 4).

Fig. 4. An animated guide, Multi Mega Book in the CAVE®, Franz Fischnaller et al., 1997.

In some cases, the limitations presented in the development of intelligent characters for virtual worlds are overcome by the use of avatars or actors, that is, the virtual representations of real people [4]. In the *NICE* project [14], an educational VR environment where children could collaboratively plant a garden and construct stories about their activities, intelligent agents were originally conceived to act as mentors, by helping the students to complete tasks, as well as characters to progress a story. In NICE, the construction of the environment is designed to foster collaboration between remotely located users. Through the use of avatars, geographically separated learners are simultaneously present in the virtual environment (Fig. 6). This ability to connect with learners at distant locations, enhanced by visual, gestural, and verbal interaction was employed to develop unique collaborative experiences for both the students and the educators.

Initial research indicated that current technical developments were not advanced enough to construct "intelligent" agents that could respond to the needs of students from different locations and suspend disbelief, even though the final stories produced were not complex. By replacing the agents with avatars of (real) people, teachers or parents participated, either as members of the groups, or disguised as characters in the environment. This allowed teachers to mentor the children in person, to guide parts of the activity from «behind the scenes» and to help shape more interesting and engaging stories [13].

Fig. 5. Children interacting with avatars in a virtual garden to construct stories. The NICE project, Maria Roussos et al. http://www.evl.uic.edu/tile/NICE/

5 Concluding Remarks

The world of interactive narrative is challenging all preconceived notions of the art of storytelling. Traditional narrative patterns where a story is defined as a -mostly linear-series of interrelated events in a setting with scenery, props and actors, seems largely unsuitable for a virtual environment where exploration of an environment formed around the user defines the more dominant model.

The questions evoked by this limited review of virtual reality endeavors that draw on the powers of storytelling, reveal that before one can speak of narrative in virtual worlds, a whole new mindset of use and a whole set of tools must be developed. A mindset that takes into account immersion and interactivity as conflicting but also additive features to storytelling, that explores the role of the user/viewer/visitor/participant as author, narrator, or an essential part of the narrative experience, and that regards aesthetic form, representation, emotional involvement, and content as interconnected.

References

1. Anstey, J., Pape, D., and Sandin, D.: Building a VR Narrative. Proceedings of SPIE Vol. 3957 Stereoscopic Displays and Virtual Reality Systems VII (2000).
2. Bates J.: The Nature of Characters in Interactive Worlds and The Oz Project. Carnegie Mellon University Technical Report, CMU-CS-92-200 (1992).
3. Benayoun, M.: World Skin. In catalog of Ars Electronica Festival 98, Linz, Austria (199).
4. Cooper, M. and Benjamin, I.: Actors, Performance and Drama in Virtual Worlds. In Computer Graphics: Developments in Virtual Environments (1995) 255-265.
5. Fischnaller, F. and Singh, Y.: Multi Mega Book. In catalog of Ars Electronica Festival 97, Linz, Austria (1997).

6. Gaitatzes, A., Christopoulos, D., Voulgari, A. and Roussou, M.: Hellenic Cultural Heritage through Immersive Virtual Archaeology. Proceedings of 6th International Conference on Virtual Systems and Multimedia, Japan (2000).
7. Laurel B.: Computers as Theatre. Addison-Wesley Publishing Co. (1993).
8. Lebowitz, M. Creating Characters in a Storytelling Universe. In *Poetics* **13** (1984) 171--194.
9. Malone, T.W and Lepper, M.R.: Making Learning Fun: a Taxonomy of Intrinsic Motivations for Learning. In: Snow R. and Farr M. (eds.): Aptitude, Learning, and Instruction: Cognitive and Affective Process Analyses, Lawrence Erlbaum Ass. (1987).
10. Pascoe, D.: Peter Greenaway: Museums and Moving Images. London: Reaktion Books, Ltd. 1997).
11. Rokeby, D: The Construction of Experience: Interface as Content. In Digital Illusion: Entertaining the Future with High Technology, Dodsworth Jr., C. (Ed.). Addison-Wesley Publishing Co. (1998) 27-47.
12. Roussos, M. and Bizri, H.M.: Mitologies: Medieval Labyrinth Narratives in Virtual Reality. Proceedings of 1st International Conference on Virtual Worlds, France (1998) 373--383.
13. Roussos, M., Johnson, A., Moher, T., Leigh, J., Vasilakis, C. Barnes, C.: Learning and Building Together in an Immersive Virtual World. In *Presence* Journal 8(3), MIT Press (1999) 247-263.
14. Roussos, M., Johnson, A., Leigh, J., Vasilakis, C., Barnes, C., and Moher, T.: NICE: Combining Constructionism, Narrative, and Collaboration in a Virtual Learning Environment. Computer Graphics, 31(3), New York. ACM SIGGRAPH (1997) 62-63.
15. Ryan, M.: Narrative as Virtual Reality. The Johns Hopkins University Press (2000).
16. Steiner, K. and Moher, T.G.: Scaffolding Story Construction with Interactive Multimedia. Journal of Educational Multimedia and Hypermedia *3(2)* (1994) 173-196.

Virtual Storytelling of Cooperative Activities in a Theatre of Work

Leonie Schäfer[1], Wolfgang Prinz[1], Amanda Oldroyd[2], and Lesley Gavin[3]

[1] GMD - German National Research Center for Information Technology
Schloss Birlinghoven
D-53754 Sankt Augustin, Germany
{leonie.schaefer, wolfgang.prinz}@gmd.de

[2] BTexact Technologies
Adastral Park, Martlesham Heath
Ipswich IP5 3RE, UK
amanda.oldroyd@bt.com

[3] VR Centre for the Built Environment
University College London
Gower Street
London W1CE 6BT, UK
l.gavin@ucl.ac.uk

Abstract. DocuDrama offers the generation of interactive narratives which are based on activities in a collaborative virtual environment. DocuDrama develops a model for the creation and enactment of narratives derived from the history of documents and interactions between people. It investigates how a narrative can be constructed from this information in a way appropriate for both the intended audience and the message to be conveyed. DocuDrama offers a choice of replay options which depend on the user's situation and preferences. We apply this approach within TOWER, a Theatre of Work, which allows project members to be aware of project relevant activities as well as to establish social relationships to intensify team coherence.

1 Introduction

In current work environments, people are working together in teams geographically dispersed and located all around the world. Team members communicate by telephone, email and by using information technology like Internet and shared workspaces. This form of communication supports the functional part of the project work and coordination. However, in traditional work settings where project teams are working together in the same environment a huge amount of information is passed by interpersonal communication.

This kind of informal communication and coordination plays a significant role for the successful cooperation in teams, as important factors in local cooperation are accidental meetings and peripheral awareness of ongoing activities. These factors often

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 191–200, 2001.

have a significant influence on the individual orientation in the overall cooperation process. However, current cooperation support technology does not adequately provide means for social encounters and awareness of other team members.

Another aspect of cooperative work that needs further attention is the problem of catching up with past activities after a period of absence. Often this is handled by asking a colleague or, more painfully, by retrieving history information from email communication or shared information spaces. More advanced technological support for the user, e.g. an adaptive report on past cooperative activities, is still an open research topic.

There is a need, then, for a new type of awareness infrastructure that will deliver information about the status of work, inform about ongoing activities, enable social encounters and furthermore to report on past activities. The Theatre of Work Enabling Relationships (TOWER), is a system that addresses these issues through the provision of a 3D collaborative environment. The 3D scenery of TOWER consists of a landscape that is generated from the information shared by a team. Within this information landscape the team members are represented by avatars that perform symbolic actions based on the actual activities of the respective team members. These symbolic actions are played out automatically, i.e. a user does not have to navigate her avatar through the 3D scenery, but the avatar is automatically routed to the place that represents the information, e.g. a document the user is currently working with. Depending on the actual operations performed by the user, e.g. read, create, modify, the avatar performs different symbolic gestures that indicate these operations.

By projecting this interface into the office environment, TOWER offers a stage for social encounters and that tells a story about the work process in teams and the current and past activities in a cooperative environment. Users working at different sites can see what is going on by looking at the TOWER scenery. This scenario is illustrated by figure 1.

Fig. 1. Integration of the Theatre of Work into a work setting

In order to achieve this functionality, the TOWER system is composed by a number of interacting components. Figure 2 illustrates the overall TOWER architecture.

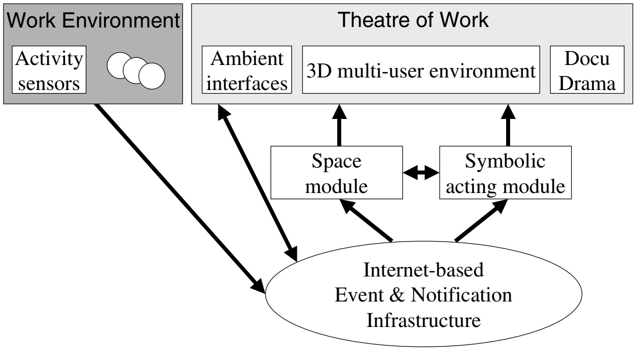


Fig. 2. Overview of the TOWER architecture

Information about current activities is collected by a number of different activity sensors that capture and recognize user activities in a real and virtual work environment and submitted as appropriate events. These sensors provide the real world data on which the story in the Theatre of Work is created. They forward this information by means of events to an Internet-based event & notification infrastructure that stores and forwards these events to interested and authorized users [2].

Two clients of this infrastructure are the space module and the symbolic acting module. The space module dynamically creates 3D spaces from virtual information environments, e.g. shared information workspaces such as Lotus Notes or BSCW, and adapts existing spaces to the actual usage and work behaviour of the users that populate these spaces. The generation of these spaces is based on the space syntax [1].

The symbolic acting module transforms event notifications about user actions into symbolic actions, i.e. animated gestures of the avatars that represent users and their activities in the environment. The 3D multi-user environment interoperates with the symbolic acting and space module for visualisation and interaction. In this component the story is actually visualised and presented to the users. The 3D visualisation is complemented by ambient interfaces integrated into the physical workplace providing activity visualisation methods beyond those of the standard desktop.

The DocuDrama component transforms sequences of event notifications and history information into a narrative of the past cooperative activities. In the remainder of this paper we focus on the description of the DocuDrama component.

2 DocuDrama – Telling a Story of Past Cooperative Activities

For effective collaborative working it is vital for teams to be able to access records of decisions made, minutes of meetings and document histories. It is also vital that new members of teams are able to catch up with what has happened in order to get a clear picture of the state of a project. Whilst many systems are available for recording changes and amendments to documents, and minutes are written recording decisions and actions at meetings, the information gleaned from these sources can be very sketchy. It can also be very difficult for team members to fully understand the context in which decisions were made or documents changed. For full understanding of what has happened it is necessary to perceive the course of events including the activities of actors [3],[4].

DocuDrama as a feature of the Theatre of Work focuses on the recording and replay of events. Activities for recording include meetings or monitoring places with a high density in avatar interaction, but also the evolving information landscape. Subject of recording in this case are information objects in the landscape and their changes over a period of time. The replay of events takes place on the user's demand. The replay is influenced by the user's interest which includes his current situation and also his personal interest profile. The following scenario illustrates one possible use of DocuDrama in the TOWER context.

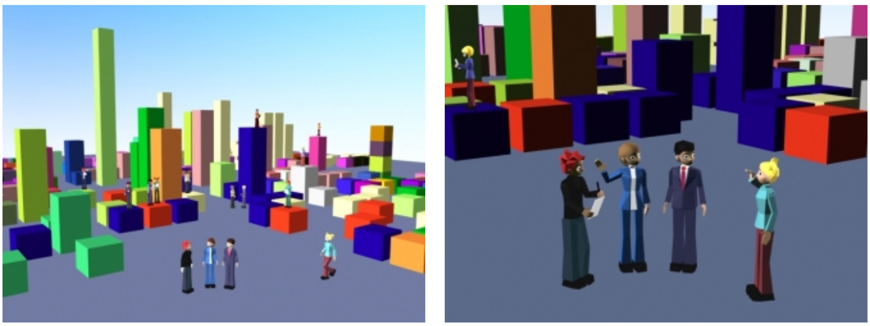


Fig. 3 a,b. Scenes from TOWER

Our TOWER World user, Alison, has been away for a week at a conference. She needs to get a quick overview of the management developments, e.g. if there have been important meetings and what have been the major topics. She clicks on the DocuDrama symbol in the TOWER portal, chooses her preferences and watches the unfolding of her personal DocuDrama story. DocuDrama stories are animation sequences which consist of a compilation of screenshots, animations and audio recordings showing important activities. Alison's particular story shows a sequence of formal and spontaneous meetings. Such sequences first display an overview screenshot of the information landscape (see Fig.3a) to show the context in which the meeting has taken place. At a deeper level the following screenshot presents the participants of the meeting as avatars in TOWER World as their meeting environment (see Fig 3b). The final screenshot in a single meeting sequence gives an overview of the documents related to this meeting, e.g. the agenda or a protocol of the meeting. The development of DocuDrama encompasses a wide range of research areas, which are discussed in the following.

Summarization is a keyword relevant to DocuDrama. It has to be decided which events are suitable for recording, on the one hand to avoid server overload, on the other to guarantee an exciting narrative. Only events should be recorded which are of interest to the user, e.g. match his interest profile. Methods to be applied to scan and aggregate this wealth of information originate from the field of summarization[16], but also include analysis and evaluation of user behavior in shared workspaces.

The virtual environment in the Theatre of Work visualizes a virtual team's shared workspace as an information landscape. This enables the user to get an overview of the project's current state, as well as to see at a glance the latest progress in the project's workflow. The composition of the information landscape plays an important role in orientation and finding the way in the virtual environment [5]. Automatic Camera Control in the virtual environment is another important factor. DocuDrama should be regarded as a kind of theatre. The user can watch passively how the narrative unfolds, with the chance to interact if he wants to. Automatic camera control guides the user through the virtual landscape and offers him the most advantageous view on the scenery. This field of work relies on experiences in movie-making and cinematography [6][7][8].

Presentation methods and the selected time period define the way the story is generated. The simple replay of event data might cover all activities which have taken place during the selected time period, but all events are given the same importance. To generate a replay of narrative quality principles of screenplay writing and interactive cinema should apply to the automated story generation [9][10].

There are several scenarios in which DocuDrama will be a useful feature for the replay of past activities. The presentation method differs depending on the current situation and preferences of the user. Several presentation methods for the story might apply which range from snapshots, movie clips to the replay of events in the current virtual environment [11][12][13][14].

2.1 Related Work

DocuDrama focuses on the recording and replay of events in a collaborative virtual environment. Related work, as discussed in the following, investigates certain features of DocuDrama, but no approach is known which uses a DocuDrama combination of research areas.

Temporal Links [11] introduces the idea of a flexible mechanism for replaying past or recent recordings of virtual environments within other virtual environments. Temporal Links is concerned with time, spatial and presentational relationships between the environment and the recording. Where Temporal Links focuses on replaying the past and its implications with the current environment, DocuDrama is concerned with selection and aggregation of history events and their replay depending on the user's scenario.

Brooks, [6][9] with Agent Stories, has investigated a model for the computational generation of narratives. This model splits the task into: defining an abstract narrative structure, collecting material and defining a navigational strategy. While Brooks offers a story design and presentation environment for non-linear, multiple-point-of-view cinematic stories, DocuDrama focuses on the automated generation of narratives by selection and aggregation of events.

Comic Chat [14] offers the representation of online communications in form of comics. The system offers an automated approach to comic generation by the use of a default selection of gestures and expressions as well as placement and orientation of comic characters. While Comic Chat concentrates on visualization of communication, DocuDrama is concerned with the symbolic representation of user's activities in a 3D environment.
Finally, DocuDrama differs from all these systems in its foundation in collaborative work and cooperation awareness. DocuDrama is the only system, which combines the replay of past activities with collaborative virtual environments, cinematography and symbolic acting.

3 DocuDrama in TOWER

This chapter describes the components in TOWER which are relevant to DocuDrama, the event and notification infrastructure (ENI), replay scenarios and the storyline generation.

3.1 Events as Sources for the Storyboard

The event and notification infrastructure (ENI) together with different sensors provide the input for the construction of the story by the TOWER DocuDrama component. ENI provides a set of methods to submit and retrieve activity events. These methods are provided either as CGI-functions that can be called by a simple HTTP-request or by a JAVA API that allows for synchronous communication.

For TOWER a number of different sensors have been realised. We differentiate between software sensors and hardware sensors. Software sensors are used to recognise user activities performed with computer systems. These can be actions such as editing a document or upload and downloading a document to a shared document management system. These sensors also recognise if a user (i.e. his workstation) is online, idle or busy. The availability of new information in open information spaces, such as the WWW, is sensed by agents that observe the content of web pages. The fact that these sensors can interact with ENI by simple HTTP-calls allows the integration of sensors in almost all modern applications that provide an HTTP-interface. Therefore it was easy to incorporate sensors as part of MS-Office documents.
Hardware sensors are used to recognise real world events. In our prototype we have made use of movement and acoustic sensors to sense the presence of people in a coffee or meeting room. These sensors permit the creation of stories that combine real and virtual activities.

Typical event data consists of: sensor-type, event-type, producer of the event, artefact in use, performed operation, date/time, expiration date, and the access control list. This attribute list can easily be extended for special application purposes since ENI does not require a special event format or registration of event schema. Events need to comply to a predefined event syntax only. This make the infrastructure very flexible since new applications can submit new event types without any administrative overhead. Further features of ENI are access control on events and reciprocity methods, i.e. a user can ask the system who is interested in events produced by oneself.

ENI provides event information via two different interfaces. A HTTP-based query interface provides methods for the retrieval of events based on event attributes. In addition, a client process can register an interest profile that consists of one or more event queries. Whenever a new event is received by ENI which matches an interest profile this event is forwarded to the appropriate client.
Beyond the retrieval and distribution function, ENI provides methods to aggregate events. This allows the aggregation and collection of certain events to a meta-event with a more expressive semantic, e.g. multiple document edit operations are combined to a single edit operation, or a certain sequence of different event types are combined to a single event. These methods are used by DocuDrama to create more expressive replays that omit repeating events. The detection of event sequences by a follow-up method is used by the camera agent to identify interaction sequences and to control the camera position accordingly.

3.2 Replay Scenarios

Presentations methods for viewing a DocuDrama story differ depending on the current situation and preferences of the user. The process of selecting and recording events within the TOWER environment can result in one or more DocuDrama narratives. Users have to select the most appropriate replay method for their needs and require a means of identifying the sequence they wish to watch first. We anticipate using snapshots and movies in addition to actual replays within the current virtual environment.

A set of snapshots from key events recorded by the camera in the virtual world will be placed in linear sequence to form a photo-story, much like images in a photograph album. These snapshots will be a range of long, mid and close-up shots, to establish the setting, the action and the characters involved in the story. The actions of avatars will be seen as though caught in mid-motion, performing actions such as reading, writing, talking with others nearby. Snapshots can be taken by the camera agent at regular intervals or at points of interactions between people or between people and objects, depending on user preference. These snapshots give the user an overview of events in a short time-frame and could be used as an interface to launch a movie clip or a replay within a virtual world. Snapshots can be mailed to the user or downloaded easily from a server.

Movie clips offer an animated view of DocuDrama. These are linear sequences that can be watched on a variety of devices and offer the ability to stop, pause and replay the story. In contrast to the snapshots, the timing of interactions will be more visible as camera movement can be included. We can create mise-en-scene shots which show how objects in the space relate to one another before cutting to selected areas of interest as specified by the user profile.

Fig. 4. A camera agent embodied as avatar

Replay within a virtual world can happen in two ways. One way is to select a time period within TOWER, e.g. 24 hours and ask for a replay of that time-period in a shortened sequence – for example 10 minutes. This is similar to time-lapse photography except that the world could be explored through manual navigation or from set viewpoints. A more structured method is to generate a set of viewpoints which take the user on a guided tour of points of interest. This guided tour allows for interaction between the user and the world, stopping to look at things from different angles or to click on documents that had been changed by a group of people. A greater period of reflection is afforded by this type of replay although the ability to view the world within the 3D browser is required.

3.3 Storyline Generation

Initial steps to generate stories from within the TOWER world focus on selecting events and watching a replay of the story in a virtual world. The user's interest profile allows the selection of events which belong to a certain topic or context. These might be events in the information landscape, avatar activities or also a user-defined selection of actions. The events belong to a set of viewpoints, i.e. an information object has two or more related viewpoints which present the object in an advantageous way. The events are now combined to a narrative as a follow-up of viewpoints. The selection of the viewpoints and their combination to a narrative in form of a viewpoint tour is performed by a camera agent (see Fig 4).

The camera agent is the main instrument of storyline generation. The camera agent selects viewpoints according to the chosen camera interest profile and combines them to a virtual tour in the 3D environment. Camera agents operate on the data of 3D objects, e.g. location, size and also position and name of relevant viewpoints, which is accessible through a database. The constantly evolving data landscape in TOWER requires on-the-fly viewpoint generation for new objects and changing selection criteria. Therefore the database contains meta data attached to clusters of objects. Context-dependent viewpoints are generated according to the selection criteria chosen. State information of the objects is stored on a MySQL database and is accessed via standard HTTP-calls. The software is written in Perl, Java and C++ respectively. The multi-user platform for TOWER is provided by the blaxxun Community Server [15].

To further enrich the narrative within DocuDrama story generation, we will introduce additional symbolic clues into the environment. These clues will represent collections of events and as such will offer additional information about the activities in the Theatre of Work. This implicit information will be represented by additional indicators such as a spotlight on an avatar or balloons rising to the sky. The status of the indicators provide a context for the user to relate individual activity to the group as a whole, whilst giving them a story-signpost or marker for points of interest in the narrative.

4 Summary and Future Work

DocuDrama stories are virtual stories based on a user's activities in a virtual collaborative environment. The stories offer a review of past events adapted to the user's interest. Events that result from cooperative activities on shared documents and information spaces provide the foundation for the construction of a story. DocuDrama offers a choice of replay options depending on the user's situation and preferences. The generation of a storyline is based on a camera agent, which selects events and combines points of interest to an interactive virtual narrative.

Our future plans are to evaluate the use of camera agents to record narrative sequences and to make the system scalable and robust. We will then investigate the types of snapshots and film clips that can be produced directly from the recording of events within the virtual environment. User input will help us select which kinds of story presentation are most suitable for a comprehensive summary over a long period of time, in comparison to a short period of time away from the system.

Another focus will be set on the presentation of history events in the current Theatre of Work. This offers interesting possibilities for interacting with the present and the past and also raises important issues relating to openness and privacy. This type of presentation is especially useful for events of the near past which still have a relation to current activities in the virtual environment.

We will also improve the options for user configuration. One option will be the selection of events relating to a user-definable context. Another option will be the definition of locations in the virtual environment which serve as observation points for monitoring events in the Theatre of Work.

Acknowledgements. We thank the members of the TOWER project team who provided helpful ideas and support during preparation of this document. The TOWER project is partly funded through the IST program IST-10846.

References

1. Penn, A., J. Desyllas and L. Vaughan, "The Space of Innovation: Interaction and Communication in the Work Environment, " Environment and Planning, 1999, Vol. 26, No. pp. 193-218.
2. Prinz, W., "NESSIE: An Awareness Environment for Cooperative Settings" in Proc. of ECSCW'99: Sixth Conference on Computer Supported Cooperative Work, S. Bødker, M. Kyng, and K. Schmidt eds., Copenhagen, Kluwer Academic Publishers, 1999, pp. 391-410.
3. Prussak, L., Knowledge in Organizations, Butterworth-Heinemann, Oxford, 1997.
4. Swan, J., S. Newell, H. Scarbrough and D. Hislop, "Knowledge management and innovation: networks and networking, " Journal of Knowledge Management, 1999, Vol. 3, No. 4, pp. 262-275.
5. Norman G.Vinson, "Design guidelines for landmarks to support navigation in virtual environments" , Proceedings of CHI 99, May 15 - 20, 1999, Pittsburgh, PA USA
6. Kevin M.Brooks, "Do Story Agents Use Rocking Chairs", Proceedings of the fourth ACM international conference on Multimedia, November 18 - 22, 1996, Boston, United States
7. B.Tomlinson, B. Blumberg, D. Nain, "Expressive Autonomous Cinematography for IVEs", In: Proceedings of Autonomous Agents 2000
8. Drucker, S.M. „Intelligent Camera Control for Graphical Environments", PhD Dissertation, MIT Media Laboratory 1994
9. Brooks KM (1999), "Metalinear Cinematic Narrative: Theory, Process, and Tool" MIT Ph.D. Thesis.
10. "Interactive Fiction", IEEE Intelligent Systems, November/December 1998
11. Chris Greenhalgh, Jim Purbrick, Steve Benford, Mike Craven, Adam Drozd, Ian Taylor, "Temporal Links: Recording and Replaying Virtual Environments" in Proceedings of the 8th ACM international conference on Multimedia, pp. 67-74, 2000, ACM Press
12. Greenhalgh, C., Benford, S., Taylor, I., Bowers, J., Walker, G. and Wyver, J., Creating a Live Broadcast from a Virtual Environment, in Proc. ACM SIGGRAPH'99, pp.375-384, ACM Press
13. Mike Craven, Ian Taylor, Adam Drozd, Jim Purbrick, Chris Greenhalgh, Steve Benford, "Exploiting interactivity, influence, space and time to explore non-linear drama in virtual worlds" Proceedings of the SIG-CHI Conference on Human Factors in Computing Systems, 2001, Pages 30 - 37
14. David Kurlander, Tim Skelly and David Salesin; "Comic Chat"; Proceedings of the 23rd annual conference on Computer graphics, 1996, Pages 225 - 236
15. blaxxun, http://www.blaxxun.com/
16. A Bibliography of Research in Text Summarization, http://www.cs.columbia.edu/~radev/summarization/

Virtual Storytelling for Training: An Application to Fire Fighting in Industrial Environment

Ronan Querrec and Pierre Chevaillier

Laboratoire d'Informatique Industrielle
Ecole Nationale d'Ingénieurs de Brest
Technopôle Brest Iroise
CP 15 ; 29608 Brest Cedex, France
[querrec, chevaillier]@enib.fr
http://www.enib.fr/LI2

Abstract. The goal of this project is to build a virtual reality platform to educate fire fighters officers. Virtual reality allows to immerse users (teacher and learners) in a universe where the physical environment and human actors behavior are simulated. We propose an architecture where everything is an agent: reactive agents (natural phenomena), cognitive agents (firemen) and avatars (users). The two last types of agents coordinate their actions: they play a role in an organization to execute pre-established missions in team.

1 Introduction

The goal of this project is to create a Virtual Environment for Training fire fighting officers. This project aims to educate officers to manage fire fighter teams during an incident in an industrial site. Like in previous fire fighter VETs [8], we need to simulate the physical environment and immerse the user, but we need also to simulate the fire fighter teamwork. The solution we propose is to simulate every component of the VE by an autonomous agent: physicals entities like in [1], the simulated fire fighters and the avatars representing the users (learners and teacher). This article deals with the description of the three classes of agents, the implementation of the simulated fire fighters organization inspired from [3] and [4]. It focuses on the description of procedures as virtual storytelling to be played by virtual agents. In this article, the models presented are illustrated by an application: a training platform for fire fighters.

2 Agents

Three classes of agents have been defined. First, the reactive agents implement behaviors corresponding to physical and chemical phenomena, "the four elements": air, water, earth and fire. These agents represent the elements of the scene (fuel tank, vehicle) and all the tools used by the characters (fire hose nozzle...). Their behavior is based on reflexes and the computation of physical and chemical models (gas propagation and explosion).

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 201–204, 2001.

The second class, the cognitive agents, are advanced reactive agents. In the same way, they have to react to physical phenomena so they exhibit reactive behaviors. They have cognitive capacities (introspection, reasoning...). Cognitive agents represent the simulated fire fighters that are supposed to collaborate to execute a plan. We have defined a model of organization which defines the different roles of the team and defines the possible actions of each role. The cognitive agent's behavior is based on the selection of actions according to their perception of the environment and the evolution of the shared plan.

Each trainer is represented in the VE by an avatar (the third class) that can also collaborate with the cognitive agents. The model of organization we have defined allows an avatar to play one of the roles in the team, the avatar is then seen as a cognitive agent by the other agents of the organization. The principal action of a user is to order the realization of procedures.

3 Procedure Description and Collaboration

A team of cognitive agents knows a finite number of procedures (25 for firemen). Those procedures are pre-established, that means that each fire fighter knows perfectly the procedure and knows that the other member knows them also. The procedure can then be seen as the shared knowledge of the team. A procedure is a temporal organization of the actions described in the roles of the organization. That means that a domain specific procedure is translated into a set of temporal constraints. A temporal constraint is composed of a constraint (`Meet, During`...) and two terms (actions of the roles) [2]. The figure 1 shows the translation of a fire fighter procedure to temporal constraints. A team of agents has a constraints manager to verify the execution of the procedure and to help agents to select the actions to do.

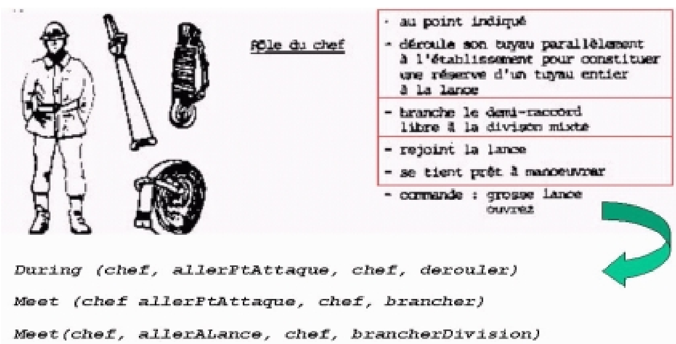


Fig. 1. Translation of a fire fighter procedure to temporal constraints

The description of the procedure can be seen as a scenario in virtual storytelling, but this scenario is played by autonomous agents in a dynamic environment. That means that the agents may encounter unpredictable situations which will force them to adapt the procedure. The main behavior of a cognitive agent is to choose the actions to do. Each action has a goal and preconditions (boolean expressions). The agent selects the actions (from his role) which have to be done according to the procedure. For all of these actions, the agent verifies if it's goal is not reached and if the precondition is satisfied. If the goal of the action is yet reached, the agent doesn't do the action. If the precondition is not verified, it searches an action of it's role that could satisfy the precondition. If it don't find it, it refers to the agent playing the role of chief which have to find a agent in the team who knows how to verify the precondition.

4 Application

The application we propose is developed with our ARéVi-oRis VR platform [5]. As a first example, a gas storage site has been modeled. The Virtual Environment is composed of gas tanks and trucks which are modeled by reactive agents. Their behavior is to compute their internal state (temperature...) from the other external entities that can modify it (fire, water...). The actions of such agents are fire, gas propagation and water jet... The different evaluation functions (internal temperature of a gas tank ...) are supplied by specialists of the domain.

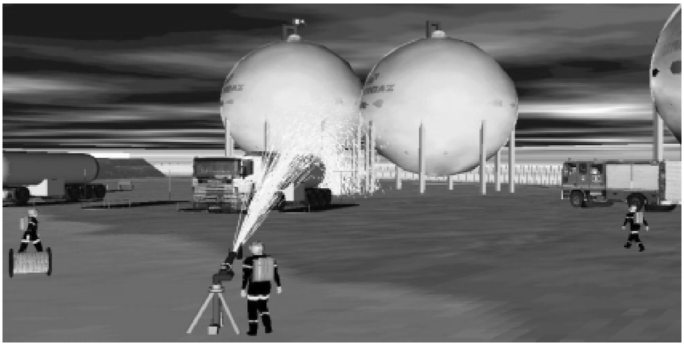


Fig. 2. A fire fighter team fighting against a leak of gas

Cognitive agents represent the fire fighters that realize the different procedures ordered by the officer (user). They collaborate following the plan of the procedure and according to the state of the environment. The procedure are provided by the fire fighters handbook. A team of fire fighters is composed of one chief which role is to fight against the incident (water a gas leak...) and two other firemen which role is to provide water to the chief (enrolling water pipe...) and help the chief. Figure 2 shows a team of firemen performing a procedure to fight against a leak of gas.

Learners are the officers which have to learn how to manage human and material resources during an incident in an industrial site. The teacher navigates in the environment and can create anomalies to evaluate the reaction of the learner.

5 Conclusion

The model we propose (Role, Team, Collaboration) for this type of simulation, permits to explain and implement the multi-agent system. This organizational model allows also the agents to reason and help the user to interact in the virtual environment (playing a role in a team for the learner and modifying the environment for the teacher). In our case, the procedures are pre-established and well known by the virtual agents, there is no planning. Those procedures can be seen as the shared knowledge of the team. By implementing explicitly this shared knowledge in the team, each virtual agent knows what to do and what the other virtual agents of it's team will need. It leads to reduce the interactions between the agents and optimizes the realization of the procedure.

The cognitive behavior of an autonomous firemen is closed to classical AI methods, we then plan to use tools like Prolog or SOAR like in STEVE [7]. The behavior of such autonomous agent working in a dangerous environment can be modified by emotional skills (fear, tiredness...), we will soon incorporate Fuzzy Cognitive Map to model such behavior like in [6].

References

1. Chevaillier, P., Harrouet, F., Reignier, P., Tisseau, J. Virtual Reality and Multi-agent Systems for Manufacturing System Interactive Prototyping. International Journal of Design and Innovation Research, vol. 2, n°1, 90 – 101
2. De Loor, P., Chevaillier, P. Generation of agent interactions from temporal logic specifications. 16th IMACS World Congress, 2000
3. Gutknecht, O., Ferber, J. Vers une méthodologie organisationnelle de conception de système multi-agents JFIADSMA 99, Hermès Science, 1999, 93–104
4. Hannoun, M., Boissier, O., Sichman, J., Sayettat, C. MOISE : un modèle organisationnel pour la conception de systèmes multi-agents. JFIADSMA'99, Hermès Science, 1999, 105–118
5. Harrouet, F. oRis: s'immerger par le langage pour le prototypage d'univers virtuels à base d'entités autonomes. PhD thesis, Université de Bretagne Occidentale, 2000
6. Parenthoen, M., Tisseau, J., Reignier, P., Dory, F. Agent's Perception and Charactors in virtual Worlds: Put Fuzzy Cognitive Maps To Work Proceedings of VIRC'01, 2001, 11–18
7. Rickel, J., Johnson, W. L. Animated agents for procedural training in virtual reality: perception, cognition and motor control Applied Artificial Intelligence, 1999, 13, 343–382
8. Tate, D. L., Sibert, L., King, T. Virtual Environments for Shipboard Firefighting Training Virtual Reality Annual International Symposium, 1997

Computer Animation and Virtual Reality for Live Art Performance

Jean-Pierre Jessel[1], Céline Jaspart[2], and Jean-Jacques Flores[2]

[1] IRIT-UPS 118 route de Narbonne, 31062 Toulouse, France
jessel@irit.fr
[2] Animaçào, 14, Rue St Bertrand, 31500 Toulouse, France
jaspart@worldnet.fr

Abstract. We present a virtual puppet performance including an animation system for real-time character animation based on motion capture. This performance may mix puppets, dance, comedy, and achieves interactivity between the manipulators and the other participants (real/virtual) on the stage. Most often, the show is aimed to children, but more complex or elaborated stories or concepts can be proposed to an older audience. The objective is to explore these new possibilities with low-end hardware and devices. For the future, including new concepts such as behavioral simulation would lead to animate directly main characters whereas the other ones would be autonomous.

1 Introduction

We present a collaboration of the Image Synthesis and Virtual Reality Group in IRIT with artists, within the framework of spectacles of alive art, being able to mix puppets, dance, comedy and in which motion capture devices are used. The artistic approach does not propose active interactions with the public: the interactivity is reserved for the manipulators and to the other participants (real/virtual) on the stage.

We have collaborated for several years with the puppeteers of Animaçào, to use computer animation in live art performance. The increase in the use of motion capture devices caused a rebirth of the use of puppets for computer animation and for the creation of movement to be associated to offline character animations, and also for real time animation. The motion capture allows also the play of the animators with other characters, actors, dancers, or simply with themselves, during the animation of abstract forms. The techniques of capture thus open new horizons for animation and the art of performance. The objective here is to explore these new possibilities with low-end hardware and devices, for few manipulators (and few sensors), as for the computing power (a PC with a powerful graphics board).

2 Theatrical Creation and Reality Virtual Technologies

The performance is a puppet show, or mimes, on which the Polhemus sensors are laid out. In a general way, a manipulator fit out with sensors evolves/moves at sight (as shown in figure 1.c and 1.d) or hidden (as shown in figure 1.a and 1.b), while the

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 205–207, 2001.

sensors collect the positions and orientations of its body or its hands and sometime a real puppet motion. A virtual character, or an abstract form, is animated using the data provided by the sensors. The sensor positions and orientations make it possible to reflect the gestures or the movements of the real puppet or of the animators to control one or more forms.

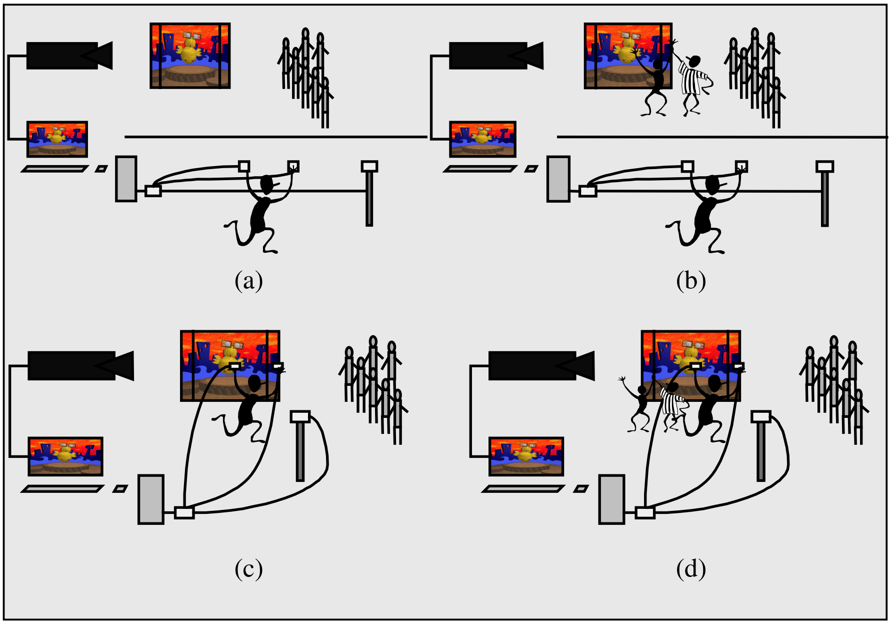


Fig. 1. Some classical installations

Interactivity is thus possible between a traditional puppet or an actors and virtual puppets or forms. A dialogue is established between the synthetic actors and the real actors, as shown in figure 1.b and 1.d, or with the animator himself, as shown in figure 1.c and 1.d. The virtual puppet can reflect the real puppet or can be opposed to the puppeteer, his behavior can also evolve. For each handling, the addition of music makes it possible to add choreography to carry out a complete spectacle. There is complementarity between the real spectacle and the video projection of the synthetic animation; people have the possibility to get the feedback from the both and to have two possible readings of the stage. Moreover, the combination of virtuality and reality and the projection of a movement from a space into another one with different aesthetic aspects offers multiple scenic possibilities, as well in the structuring as on the contrary in the improvisation, thanks to the real time.

3 Character Motion

For the characters, the model of animation used is a mixture of restitution of movement and kinematics procedural animation. The Polhemus sensors are used to print the movement of all the character or of one part only: members, eyes, nose, mouth, either directly, or after having undergone an interpretation: change of rotation axis, change of amplitude of the movement... Other parts of character are animated by predetermined movements, according to a fixed trajectory whose parameters can be changed interactively. In the same way, the keyboard can be used to swap between different elements or to modify the size, the position, the color and the texture. A data glove can also be used for the movements of the face or the members.

For example, figure 2 shows two characters. Doggy is animated with two Polhemus sensors associated to the head and to the body. Birdie is also animated with two sensors, but with several possibilities: the first one is similar to Doggy’s one, the second one uses one sensor for the whole character and another one either for the eye orientation, either for the mouth. The mouth width can also be set by keyboard interaction. Birdie’s wings are animated with a periodic movement which frequency can be set interactively.

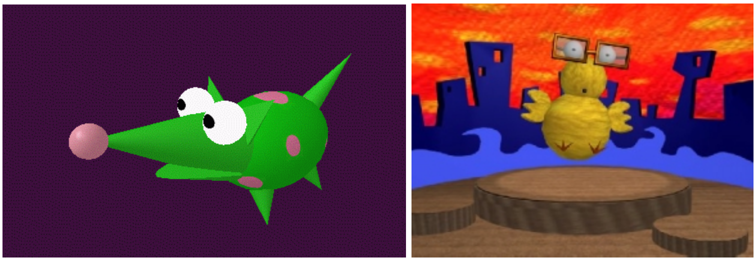


Fig. 2. Two characters: Doggy and Birdie

4 Future Developments

The animation of characters, humanoids or animals, is a difficult problem, which requires the use of complex techniques. The virtual characters are often modeled using articulated structures, which more or less coarsely model the skeleton of the character. For the design of systems of animation of characters, it is necessary to work out abstraction (motion control methods) and to consider the possibilities of interaction. The motion control methods currently used are geometrical and kinematics methods, next dynamics and higher-level methods will be implemented.

We have developed an animation platform including several motion control methods (key framing, motion capture, direct kinematics, dynamics), a script language and behavioral control, but all these works has not been set yet to the virtual puppets applications. With the dynamics, it would be possible to bind several parts of characters to produce complex movements without needing to code them or to control them specifically.

Behavioral animation would allow moving automatically the characters present in a virtual world. Behavioral simulation is studied by another part of our Group and consists of an automatic research of the behavior according to a goal to reach. For that the virtual entities have a system of evolution governing their behavior: an entity will be able to thus learn how to react to an evolving situation. By including these concepts, it would be possible to animate directly main characters whereas the other ones would be controlled themselves by their autonomous behavior.

Virtual House of European Culture: e-AGORA (Electronic Arts for Geographically Open Real Audience)

Jaroslav Adamec[1], Jiri Cizek[1], Michal Masa[1], Pascal Silondi[2], Pavel Smetana[2], and Jiri Zara[1]

[1] Department of Computer Science, Czech Technical University, Karlovo nam. 13, 121 35, Prague 2, Czech Republic
[2]CIANT - Centre International d'Art et de Nouvelles Technologies Kubelikova 27, 130 00, Prague 3, Czech Republic

Abstract. This co-operative project links up important European centres of art and culture and creates an open network with an innovative application of the new information technologies: the creation of an unconventional and highly efficient communication base founded on an interactive, multimedia and trans-disciplinary approach to the production and presentation of contemporary forms of performing arts. The main channel of this project is the revolutionary communication and navigation system e-AGORA, which enables visitors to move intuitively in a virtual 3D environment on the Internet. In the VIRTUAL HOUSE OF EUROPEAN CULTURE a broad spectrum of the public can investigate contemporary Euro-regional artistic programs in real time and communicate interactively. At the same time, it is an instrument of individual and collective artwork in the domain of contemporary performing arts as it opens up a new horizon for a multidisciplinary form of artistic expression and presentation of artworks. The implementation of the project VIRTUAL HOUSE OF EUROPEAN CULTURE also has important theoretical and educational dimensions: a series of practical workshops, an international academic conference, thematic exhibitions, the production of the e-AGORA CD-ROM and a printed publication.

1 Objectives

The complex project **VIRTUAL HOUSE OF EUROPEAN CULTURE** has arisen as an active and open network of several important cultural centres; their previous co-operation and experiences have resulted in a revision of the old communication models. In the future, new types of co-operation and new instruments of communication will answer the need for a more efficient regional exchange; these will be used for the first time in the project **VIRTUAL HOUSE OF EUROPEAN CULTURE**. This common house stands on the firm foundation of actual artistic exchange between cultural centres of similar orientation.

It will introduce, however, the revolutionary feature of a common virtual space: **e-AGORA**. This will be shared not only by the connected cultural centres, but also by the artists and the general European public. The internet platform is open practically to everyone, without distinction: if offers **multi-participant** and **multi-lingual communication** and **interactive cultural entertainment**.

O. Balet, G. Subsol, and P. Torguet (Eds.): ICVS 2001, LNCS 2197, pp. 208–211, 2001.

The multimedia navigation system **e-AGORA** will introduce everyday actual information (sound, music, image, video) from the existing spaces of five European cultural centres using the most modern information technologies. This information will be presented in real time in the virtual spaces of e-AGORA on the Internet. Visitors to this virtual 3D space will have the opportunity to select their own individual avatars; these avatars will represent them and enable them to communicate interactively with other avatars. The passive viewers of the general public (contemporary TV) will become individual members of a virtual European community, which is a revolutionary alternative to the local and mass-media limited approach to culture, art and information. **e-AGORA** will make use of the actual interiors of the connected cultural centres (e.g. DE WAAG in The Netherlands, Palace Akropolis in Prague etc.) for the **modelling of the virtual 3D space**.

When it is fully operative, the **VIRTUAL HOUSE OF EUROPEAN CULTURE** will be a unique site for artwork and the reception of artistic programs across the European continent. At the same time, however, it will remain an **open structure** with possibilities for further **expansion** and **the integration of cultural centres in other European countries**.

The transfer of artistic and technological information on a European scale will be so inventive, thanks to the key channel **e-AGORA**, that it will significantly influence the actual regional cultural exchange and prepare a new instrument for independent multidisciplinary artwork for creative artists in the field of the performing arts.

2 E-Agora Architecture

E-Agora is a multi-user virtual environment aimed mainly at social interaction. For such a system two main decisions have to be made prior to the implementation: how to render the 3D scene and how to communicate between participants' machines.

Since we wanted to spare time and develop the system at minimum cost, we decided to exploit existing technologies to the most extent possible. Thus, instead of implementing our own rendering engine, we chose to base our system on VRML and to adopt one of the existing and freely available rendering engines – VRML browsers.

For the same reasons, to support communication between participants' machines, we used an existing Java library (DILEWA/GV [1, 2]) that is being developed by our research group. The library deals with distribution of messages among several machines connected to the Internet and solves the problem of bringing later connected users (latecomers) up-to-date on the current state. The communication pattern is based on the client-server model.

A typical implementation of a networked virtual environment has to consider the following issues [3]: a shared sense of space (participants have the illusion of being located in the same space), a shared sense of presence (participants perceive each other by the help of avatars), a shared sense of time (real time interaction with the world), a way to communicate (chat, gestures, voice, video), a way to share (the environment is shared, every change is visible to all participants).

The following text explains our approach to implement these features in E-Agora system.

2.1 E-Agora Client

The client consists of VRML browser (responsible for rendering the scene) and Java applet (responsible for communication issues and the scene control). The browser runs as a plug-in of Internet browser, which accomplishes the delivery of VRML files to the plug-in.

The VRML browser renders the scene composed of a shared environment, participants' avatars and control components. We have decided to incorporate control components (for example gesture selection panel or chat-board) to the scene for two reasons. First, we wanted to provide the users with a pure 3D interface, making the view of the application consistent. Second, we wanted to stay within VRML to ensure easy portability of the system.

The connection to the server is maintained by a Java applet encapsulated in a VRML Script node. To support basic features of the MUDVR, following information is distributed among clients: notifications when a user enters/leaves the system, specifications of the users' avatars (URL of the VRML file), positions and orientations of the avatars, identifications of the gestures being performed, chat strings and environment changes. The DILEWA/GV library has been used to represent and distribute these data and the details will be discussed in the next section.

To control and receive the response from the scene, the applet is connected via VRML routes with the dynamic entities in the scene (avatars, control components and dynamic parts of the environment). These entities can generate events as a response to user's interaction (events are passed to the applet) and/or their state should be modified by the applet accordingly to the information received from the server (events are passed to the entities).

For example, clicking on another user's avatar brings up a chat-board and the user can type a message. In the background the avatar generates an event, which is handled by the applet. The applet determines the recipient and brings up the chat-board by sending another event to chat-board component. When the user clicks OK button on the chat-board, the message is sent to the recipient. Again, in the background, the chat-board generates an event containing the message and closes itself. The event is processed by the applet that communicates the message to the recipient (through the server). When the recipient's client receives the message, it sends an event containing the message to the chat-board component – the chat-board on the recipient's client is brought up with the message shown.

2.2 E-Agora Server

As we have seen in the preceding section, various information has to be exchanged among clients (notifications, gesture identifications, avatar specifications, positions, orientations…). Moreover, since clients always load the original VRML scene that is unaffected by later changes, the system should also bring latecomers (users connected later to the system) up-to-date on the current state. For example, later connected client should receive information concerning all previously connected clients to display their avatars in appropriate positions.

We chose to exploit DILEWA/GV server, which was designed especially for such purposes. It allows creation and distribution of so called *general variables*.

General Variable consists of a name for its unique identification and a list of *commands* performed on the variable. A typical command sets the variable to an arbitrary *value*. The flexibility of the concept is based on the fact that the value can be compounded of any number of any primitive data types. It can be a simple value as well as a heterogeneous structure. When a user attempts to interact with the world (navigate through the world, click on another avatar, perform gestures…), the client application creates adequate variable and adds a specific command containing a value representing the user's action. The variable is then sent to the server, which is responsible for broadcasting the variable to other clients. Finally, the receiving client should decode the meaning of the variable and replay the original action locally. Additionally, the server stores all variables in a journal, which could be sent to latecomers to update their state. A set of flags associated with every variable controls its distribution and storage; it determines whether the variable should be sent to all connected clients or to subset only and whether the variable should be stored and how. Three storage methods are provided: not stored variables (distributed only), persistent variables and temporary variables (deleted from the server as their creator disconnects).

Let us illustrate the use of general variables in the E-Agora system with three examples. In the first example there is a variable containing information about the user (avatar URL and nickname). This variable is sent to all clients and stored temporarily at the server until the user disconnects from the system. In the second example the variable represents user's gestures. It is also sent to all clients, but it is not stored at the server, since latecomers are typically not interested in gestures performed prior to their connection. In the last example a variable represents a chat string. For the same reason, the variable is not stored at the server, too. In contrast of the previous example, the variable is not sent to all connected clients, but to the recipient only.

3 Future Work

Our future effort will be aimed at making the system more stable and scalable by implementing UDP protocol in addition to TCP. This can be accomplished by an additional variable flag that will determine the reliability of the distribution. Next, we plan to add more shared dynamics to the environment (light switches, doors, desk games…). Since the client provides limited support of *NetworkNodes* as proposed in [4], this can be done by an integration of specially designed VRML objects with the environment.

References

1. Masa M., Zara J.: The Distributed Learning Environment Without Avatars, Proceedings of Information Visualization 2000, IEEE, London (2000)
2. Masa M., Zara J.: The General Variables Concept: A Simple Step from Single- to Multi-user Environment, Proceedings of Information Visualization 2001, IEEE, London (2001)
3. Singhal S., Zyda M.: Networked Virtual Environments, ACM Press, New York (2000)
4. Living Worlds, http://www.vrml.org/WorkingGroups/living-worlds/

Author Index