Gérard Subsol (Ed.)

LNCS 3805

Virtual Storytelling

Using Virtual Reality Technologies for Storytelling

Third International Conference, ICVS 2005
Strasbourg, France, November/December 2005
Proceedings

Springer

Lecture Notes in Computer Science 3805

Gérard Subsol (Ed.)

Virtual Storytelling

Using Virtual Reality Technologies for Storytelling

Third International Conference, ICVS 2005
Strasbourg, France, November 30 – December 2, 2005
Proceedings

Springer

Volume Editor

Gérard Subsol
81 chemin de la Garenne
34160 Saint Geniès des Mourgues, France
E-mail: gerard.subsol@wanadoo.fr

Library of Congress Control Number: 2005936338

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

ISSN 0302-9743
ISBN-10 3-540-30511-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-30511-8 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springeronline.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11590361 06/3142 5 4 3 2 1 0

Preface

The 1st International Conference on Virtual Storytelling took place on September 27–28, 2001, in Avignon (France) in the prestigious Popes' Palace. Despite the tragic events of September 11 that led to some last-minute cancellations, nearly 100 people from 14 different countries attended the 4 invited lectures given by international experts, the 13 scientific talks and the 6 scientific demonstrations.

Virtual Storytelling 2003 was held on November 20–21, 2003, in Toulouse (France) in the Modern and Contemporary Art Museum "Les Abattoirs." One hundred people from 17 different countries attended the conference composed of 3 invited lectures, 16 scientific talks and 11 posters/demonstrations.

Since autumn 2003, there has been strong collaboration between the two major virtual/digital storytelling conference series in Europe: Virtual Storytelling and TIDSE (Technologies for Interactive Digital Storytelling and Entertainment). Thus the conference chairs of TIDSE and Virtual Storytelling decided to establish a 2 year turnover for both conferences and to join the respective organizers in the committees.

For the third edition of Virtual Storytelling, the Organization Committee chose to extend the conference to 3 days so that more research work and applications could be be presented, to renew the Scientific and Application Board, to open the conference to new research or artistic communities, and to call for the submission of full papers and no longer only abstracts so as to make a higher-level selection.

We hope that all the objectives were met: Virtual Storytelling 2005 comprised 4 invited lectures, 21 scientific talks and 9 posters/demonstrations. Moreover, Virtual Storytelling 2005 was organized within the framework of the Virtual Reality Week of Strasbourg that also combines the Annual Workshop of AFIG, the French Association of Computer Graphics and the 2nd France-Asia Workshop in Virtual Reality. The conjunction of all these scientific meetings and the close collaboration at a European level with the DAPPPLE (Drama and Performance in Pleasurable Personal Learning Environments) research network, the TIDSE conference organizers and the INSCAPE Integrated Project made Virtual Storytelling 2005 a key high-level scientific and artistic event in 2005.

Strasbourg, November 2005 Gérard Subsol

Organization

Scientific Organization

- Gérard Subsol, Montpellier, France

Local Organization

- Piotr Szychowiak, Iconoval - Pôle Image Alsace, Strasbourg, France

Scientific Coordination

- Ruth Aylett, DAPPPLE Network, Heriot Watt University, Edinburgh, UK
- Olivier Balet, Virtual Reality Dept., C-S, Toulouse, France
- Dominique Bechmann, France-Asia Work., AFIG, Univ. of Strasbourg, France
- Stéphane Donikian, AFIG, IRISA, Rennes, France
- Stefan Göbel, TIDSE, ZGDV e.V., Darmstadt, Germany
- Jean-Pierre Jessel, AFIG, University Paul Sabatier, Toulouse, France
- Ulrike Spierling, TIDSE, FH Erfurt, University of Applied Sciences, Germany
- Alain Tubiana, Iconoval - Pôle Image Alsace, Strasbourg, France

Scientific and Application Board

- Ernest W. Adams, Designersnotebook, UK
- Espen Aarseth, Center for Computer Games Res., Univ. of Copenhagen, Denmark
- Angélica de Antonio, Universidad Politecnica de Madrid, Spain
- Clive Fencott, School of Computing, University of Teesside, UK
- Yoshifumi Kitamura, Human Interface Engineering Lab., Osaka University, Japan
- Nadia Magnenat-Thalmann, MIRALab/C.U.I., University of Geneva, Switzerland
- Ana Paiva, INESC-ID, Lisbon, Portugal
- Arghyro Paouri, Multimedia Dept., INRIA Rocquencourt, France
- Maria Roussou, UCL, London, UK & makebelieve, Athens, Greece
- Maureen Thomas, Cambridge University Moving Image Studio, UK
- Bernd Willim, The German Film School GmbH, Elstal, Germany

Program Committee

- Steffi Beckaus, Interactive Media, Univ. of Hamburg, Germany
- Erik Champion, University of Queensland, Australia
- Elliott Dumville, Dept. of Architecture, University of Cambridge, UK
- Marientina Gotsis, School of Cinema, Univ. of Southern California, USA
- HyungSeok Kim, MIRALab, University of Geneva, Switzerland
- Miguel Lozano, Lecturer, Intelligent Characters Group , Univ. of Valencia, Spain
- Narcis Pares, Universitat Pompeu Fabra, Barcelona, Spain
- Jaime Ramírez, Computer Science School, Technical University of Madrid, Spain
- Hyewon Seo, ChungNam National University, South Korea
- Richard Wages, Researcher, NOMADS Lab, Köln, Germany

Additional Reviewers

- Kristopher Blom, Fraunhofer IMK, Germany
- Janine Fron, (art)^n, Chicago, USA
- Tracy Fullerton, School of Cinema, Univ. of Southern California, USA

Table of Contents

Virtual Reality Technologies

Virtual Characters

Drama and Emotion

Interactive Digital Storytelling

New Ways of Narrative

Interactivity

Applications

Acknowledgements

The 2005 edition of Virtual Storytelling was a joint initiative of Iconoval – Pôle Image Alsace and the French Virtual Reality Group (GT-RV). It was organized in collaboration with the Annual Workshop of AFIG, the French Association of Computer Graphics and the 2nd France-Asia Workshop in Virtual Reality.

Virtual Storytelling 2005 also benefitted from the scientific support of the DAPPPLE (Drama and Performance in Pleasurable Personal Learning Environments) research network and from the TIDSE (Technologies for Interactive Digital Storytelling and Entertainment) conference organizers.

The conference organizers would like to thank Iconoval and its institutional partners — French Ministry of Research, Alsace Regional Council, General Council of Bas-Rhin, and Urban Community of Strasbourg— for funding the conference. This allowed for low registration fees in order that researchers, artists, students and SME representatives could easily attend the conference.

The special sessions “Virtual Reality Technologies for Storytelling” and “Interactivity” were partially funded by the 2nd France-Asia Workshop in Virtual Reality supported by the French Ministry of Foreign Affairs, the French National Centre for Scientific Research (CNRS) and the French Institute for Research in Computer Science and Control (INRIA), and the DAPPPLE research network jointly supported by the UK Engineering and Physical Sciences Research Council (EPSRC) and the Arts and Humanities Research Council (AHRC).

The conference website installation and management was graciously offered by the INSCAPE Integrated Project supported by the European Union.

We would also like to thank the members of the Scientific and Application Board who helped the organizers to define the conference topics and proposed names of experts for the Program Committee. The members of the Program Committee deserve special acknowledgement for their superb job in reviewing the full papers with such remarkable care and moreover... for meeting the deadline!

Finally, we wish to thank all the people who were involved in the organization of this conference on so innovative a topic.

Virtual Reality Technologies

Virtual Reality Technology and Museum Exhibit

Michitaka Hirose

Research Center for Advanced Science and Technology (RCAST),
The University of Tokyo, 4-6-1, komaba, Meguro-ku, Tokyo, Japan
hirose@cyber.rcast.u-tokyo.ac.jp
http://www.cyber.rcast.u-tokyo.ac.jp/

Abstract. Museum is now a good experimental field of VR technology. In this paper, several museum exhibits which utilized latest VR technology which include "High resolution VR" supported by sophisticated image environment such as IPT (immersive projection technology), and "real world VR" supported by mobile technology such as wearable computers are introduced. Interaction design for these kind of novel exhibits is also discussed.

1 Introduction

More than 15 years have passed since the term "Virtual Reality (VR)" was coined. This technology enabled us to step into a computer generated artificial 3D world, grasp virtual objects or enjoy various experiences which are impossible in the real world. Sometimes this pseudo-experience is more fascinating than real experience because objects that do not exist in the real world can be easily simulated. One of the most important role of this technology is to visualize various objects difficult to understand through intuitive experiences. In this context, VR technology should be able to provide various possibilities to conventional museum exhibits.

2 High Resolution VR and VR Theater for Museum

2.1 IPT (Immersive Projection Technology)

During this decade, VR technology has made significant progress. For example, IPT was invented at the University of Illinois as a sophisticated VR system. Virtual world generated by IPT has very high quality and is also very stable because the IPT consists of already commercialized components such as stereo projectors, high end graphics workstations etc.. Fig. 1 shows "CABIN" which was constructed at IML (Intelligent Modeling Laboratory) at the University of Tokyo in 1997.

This high quality environment contributes to allow us to apply VR technology to more serious applications such as a museum exhibit. In fact, recently, the use of VR technology in museum exhibit has becoming one of its most important application fields. In a virtual world, we can interact with historical architecture, artifacts and events recorded or simulated using computers.

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 3 – 11, 2005.

Fig. 1. CABIN was constructed in IML (Intelligent Modeling Laboratory) at the University of Tokyo in 1997

2.2 VR Theater for Mayan Exhibition

The “Mayan Civilization” exhibition was held at the Science Museum in Ueno for two months from 18 March through 18 May 2003 and a large-scale VR theater was constructed there in collaboration with our laboratory. The theater was an important element of the exhibition. Of the over 200,000 visitors, 120,000 visited the theater and enjoyed VR images. Reconstructed Copan ruins were projected onto a large screen, and visitors enjoyed their virtual stroll of the ruins, which actually exist on the other side of the earth. Conventionally, in most ancient civilization exhibitions, excavated articles are displayed and are simply explained. However, displays and explanations are not sufficient for most of the museum visitors to understand the location of the excavation, the extent of the ruins, and the lifestyle of the people at that time.

This might have been the first time in Japan that this scale of VR system was used for a full-fledged exhibition. The scale of the VR theater is not the sole reason why this Mayan exhibition was revolutionary. One-third of the exhibition, which was the major part, consisted of image exhibits; there was a clear distinction between this exhibition and conventional VR exhibitions, which are usually an appendage to the principal exhibition.

The theater included a huge arc-shaped screen (4 m x 14 m), three Hi-Vision equivalent images were projected on the screen (Fig.2). A large-capacity graphic workstation (Onyx, SGI) was brought in for image generation.

One significant point of using VR technology in a museum may be the use of its interactivity, which is the fundamental difference between VR and other image technologies, such as movies and TV. We have spent much effort on developing this interactivity for use in a large theater.

We planned to use the interactivity of VR during a 15 min guided tour by actors playing the roles of a king, an astronomer and an archeologist. During the guided tour, the actors operated a game-pad-type controller and gave commentary in accordance

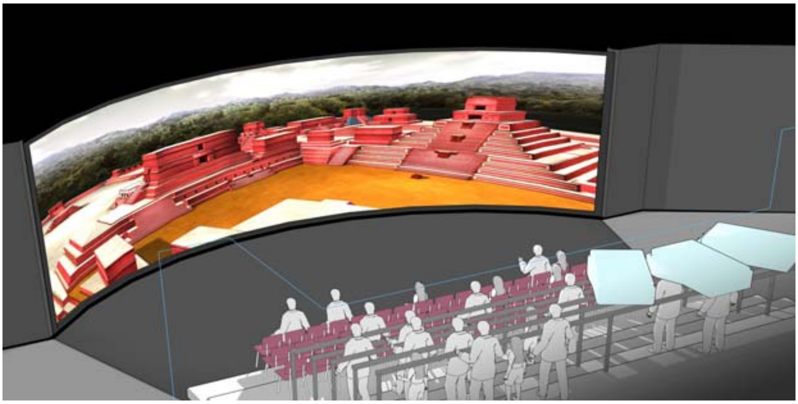


Fig. 2. Virtual reality theater was constructed for "Mayan Civilization" exhibit at the national science museum

with the story written for their roles while walking through an acropolis and tombstones. When we refer to a VR theater, we tend to consider a direct interaction between the audience and VR images. However, in this exhibition, the actors used the interactivity of images to supplement ad lib and pauses. That is, VR images are used like elements of a set in a play. This manner of using images seems to distinguish the VR image contents from regular game contents. I believe that the effectiveness of story-telling-style interactivity, which is not the same as that of games, has been proven.

The theater is designed to function on broad band network. Two PCs are installed immediately outside the theater. Information on the environment of the VR in the theater is input to the computers via network so that visitors can enjoy the tour of the ruins interactively. Visitors can freely move around in the VR world and look into places they are interested in by operating the PC controller. If the visitor stops operating the controller, the PC reverts to the view angle of the guide. That is, the relationship of the visitor with the theater and PC varies; at one time the visitors receive guidance from the theater and at another, the visitors interact of their own accord. Various ideas are being tested regarding the rules of the distribution of images using digital networks; the guidance using the mildly restricted concept presented here may be a promising system.

2.3 Virtual Mayan Ruins

Dr. Seiichi Nakamura of the Honduras National Human History Research Laboratory, supervised this VR exhibition of Mayan ruins. An acropolis with an area of approximately 350 m x 650 m has been reconstructed as a three-dimensional (3-D) world of images using computer graphics (CG). For texture mapping, 4,000 photos were used at multiple resolutions for the effective representation of the images of many tombstones existing in the ruins and trees present nearby.

Fig. 3. Virtually Synthesized No.16 Shrines at Copan ruins

Fig. 3 shows a pyramidal structure called the No. 16 shrines. The effect of texture mapping to represent the surface of crumbling stone walls was magnificent. It is said that many Mayan shrines were constructed on top of buried old shrines. Fig.4 shows one such case; we can see the inner structure of the ruins. The image shows Rosalia shrine which is present underneath the newer structure. Further down, Margarita shrine is buried.

In reality, it is very difficult to observe the entire image of buried shrines such as Rosalia and Margarita. Moreover, large expenditure is required to construct a physical replica. We can observe VR images without entailing a high cost; this seems to provide an answer to the following question: How can we observe an object when its destruction advances with observation?

Fig. 4. Virtually synthesized Rosalia shrines at Copan ruins

Fig. 5. Virtually synthesized Copan acropolis 1,200 years ago

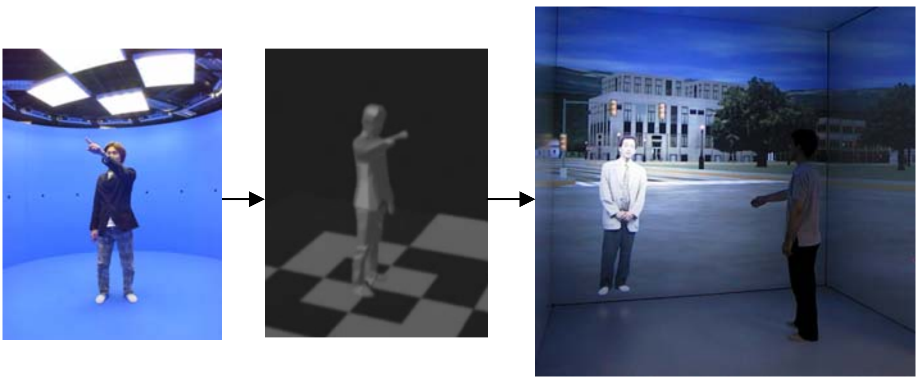


Fig. 6. 3D video avatar can be synthesized from multi-camera images

Figure 5 shows the ruins 1,200 years ago with colors reproduced on the basis of the investigation of pigments found on the ruins. Damaged buildings were reconstructed, using computers, to appear in their functional form. This reproduction has been realized as a result of the full utilization of the simulation capabilities of computers. Very accurate archeological discussion has become possible by the addition of simulation techniques to inference based on data and excavated artifacts.

To date, we have reconstructed architectural space as a physical "hardware". This is not sufficient to fully describe a civilization; the full reconstruction of ruins will be complete only when we can reproduce the life and culture of these ancient people. The important issue here is how to represent people in a VR space. So-called video avatar technology can be used here. When the world is reconstructed and the people moving about the ruins are synthesized, the breath of life is blown into the ruins. The reconstruction of "hardware" is important, but equally important is the re-creation of "software" such as human activity.

3 Real World Virtual Reality

3.1 Mixed Reality

Mixed Reality (MR) technology is also an important technology which closely relates to VR technology. One of the missions of MR technology is to import real world

Fig. 7. "Cyber-city Walker" employed IBR technique to generate realistic street scenes. Image capturing car equipped with multi cameras is shown in the left.

information into a virtual world. For example, IBR (Image Based Rendering) technology generates 3D virtual space from 2D real photographic images. This methodology contributed to improve cost efficiency in generating photo-realistic virtual world. Fig.7 is a landscape around Tokyo central station generated by a walk-through system called "Cyber-city Walker" which was a joint research project of the University of Tokyo and Canon corp.

Since by using the Cyber-city Walker, various city landscapes can be easily recorded within computers, we should be able to construct "Museum of Cities". Concept of the "Museum of Cities" is very interesting because each exhibit, a city landscape, is much bigger than a museum building , so this museum is only possible by using digital technology.

Fig. 8. Wearable computers were designed for "field exhibition "of expo 2005

3.2 Field Museum and Ubiquitous Gaming

The other aspect of MR technology is to extend virtual world to real world. By using wearable computers, we can see virtual world superimposed to real world. This technology will provide new possibility of new way of exhibit. Fig.8 shows a wearable computer prototype designed for "field exhibition" of world expo 2005. By using wearable equipments such as see-through HMD and handy GPS, we can provide various exhibits in an outdoor environment for each visitor individually. Concept of the field exhibition is fascinating because it requires basically no pavilions and one of the important themes of the expo 2005 is harmonization between artificial and natural environment.

In reality, at the expo 2005, the idea of a field exhibition has been realized as a guiding system just using a cellular phone system. Nonetheless, experiments verifying the concept of the field exhibition have been carried out on various occasions as a new framework for displays to be used in exhibitions.

Fig. 9. "Ubiquitous Gaming" was demonstrated at the national science museum

Fig.9 shows "ubiquitous gaming" demonstrated during the "TV Games and Digital Science" exhibition held at the Science Museum in Ueno from 17 July to 11 October 2004. This system can be considered to be an advanced sound guiding system. Although this system includes game-like aspects, it does not operate in a closed-image world on the tiny display of a terminal. It is a game with which visitors can enjoy their time while walking around a real space, that is, a museum.

The participants in the game use small terminals (wallstones), as shown in Fig.10. Each terminal is equipped with input-output interfaces such as a 16x16-dot LED matrix, an audio output via an earphone, a vibration output, an acceleration sensor and an azimuth sensor. When visitors stand in front of an exhibit, they can hear the voice of a game character through the earphone and his image also appears on an LED display. The azimuth sensor directs you to a specific exhibit. As another function, numerical values can be into the terminal by selecting the number of swings of the terminal, which is incorporated with an acceleration sensor.

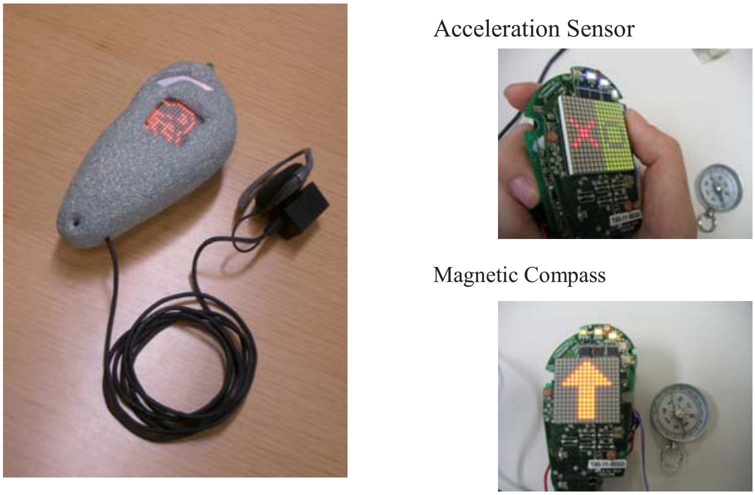


Fig. 10. Wallstones function as a mobile terminal to interface real and virtual world

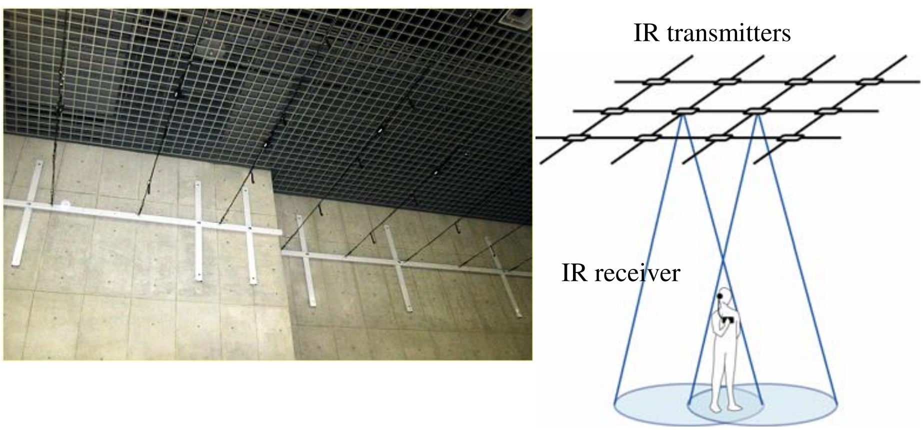


Fig. 11. More than 400 IR transmitters are used for position sensing

As shown in Fig.11, in order to support user location sensing, we installed 400 infrared transmitters arranged in the form of a matrix on the ceiling of an exhibition hall of approximately 500 m^2. A positioning accuracy of 10 cm has been achieved.

4 Conclusion

10 years ago, VR technology was a very expensive technology. However, today, most of VR hardware has become 10 to 100 times cheaper. VR technology hidden in labo-

ratories as an experimental equipment is going to various fields such as museum exhibit.

References

1. Tanikawa,T., Ando,M., Wang,Y., Yoshida, K., Yamashita, J., Kuzuoka,H., Hirose, M.: A Case Study of Museum Exhibition –Historical Learning in Copan Ruins of Mayan Civilization- . Proc. IEEE Virtual Reality conf. 2004, (2004) 257–258
2. Hiyama, A., Hirose, M.: Wearable Computers and Field Museum. Proc. HCI Interface 2003, Vol. 3. (2003) 1406-1410

A Context-Based Storytelling with a Responsive Multimedia System (RMS)*

Youngho Lee, Sejin Oh, and Woontack Woo

GIST U-VR Lab., Gwangju 500-712, S.Korea
{ylee, sejinoh, wwoo}@gist.ac.kr
http://uvr.gist.ac.kr

Abstract. In this paper, we present a Context-based Storytelling with Responsive Multimedia System (RMS). Many researches related to virtual storytelling have been presented with the advancement of multimedia technology. However, people have different abilities to understand a story according to their experience, knowledge, age, gender, etc. In the proposed approach, virtual story is unfolded by interaction between users, Multi-modal Tangible User Interface (MTUI) and Virtual Environment Manager (VEManager) with vr-UCAM (a unified Context-aware Application Model for Virtual Environments). We adopt an interactive StoryMap and tangible interfaces into MTUI with vr-UCAM such that VEManager can control the responses of virtual objects according to the user's context. Accordingly, the users can experience a personalized story since action of virtual objects, interaction level and scene of VE are reorganized and adjusted according to a user's profile and preference. To demonstrate the effectiveness of the proposed approach, we applied it to a virtual storytelling with RMS. According to the experimental results, we observed that the combination of virtual reality technology and context-aware computing could be promising technologies that enables users to experience a personalized virtual story. Therefore, we expect that the proposed approach plays an important role in virtual storytelling applications such as education, entertainment, games, etc.

1 Introduction

With the rapid advancement of hardware and software, entertainment computing industry has been popularized during the last decade. These multimedia technologies have brought huge changes in traditional storytelling. In general, books, movies, and animations follow Aristotle's Poetics which is proper for evoking feelings and emotions of readers. However, many research activities of storytelling using novel multimedia technology have presented new narrative structure which is different from traditional structure [1][2]. In the novel multimedia technology, it is necessary to balance interactivity and narration. Thus, we need to balance between story designed by an author and emergent story unfolded by participants.

There have been three kinds of approaches to achieve the interactivity of storytelling. Plot-based approach is a plot management system which is used to directly control actions of characters, leaving only the lowest level of behaviors [3]. In character-

* This work was supported in part by CTRC, and in part by MIC through RBRC at GIST.

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 12 – 21, 2005.

based approach, broad and believable autonomous characters are placed before the player, and then a story emerges from the player's interaction [4][5]. The last one is author-based approach. This approach could be modified to be of use in interactive stories. One of characteristics in previous approaches is in controlling a balance of interaction and narration. They showed a lot of interesting results, however, they did not consider a user's level of understanding, even though persons have different abilities to understand a story according to their knowledge, experience, age, gender, etc

In this paper, we present a Context-based Storytelling with Responsive Multimedia System (RMS). In the proposed approach, virtual story is unfolded by interaction between users, Multi-modal Tangible User Interface (MTUI) and Virtual Environment Manager (VEManager) with vr-UCAM (a unified Context-aware Application Model for Virtual Environments). We adopt an interactive StoryMap and tangible interfaces into MTUI with vr-UCAM such that VEManager can control the responds of virtual objects [6]. User's context, such as profile and preference, are applied to initialize whole system before beginning a story, to make an event sequence designed by an author and to select responses of virtual objects in VE.

From the proposed approach, users can interact with VE intuitively since they easily understand the use of interface. Moreover, users can experience a personalized story in which actions of virtual objects, interaction level and scene of VE are reorganized and adjusted according to a user's profile and preference. VE. Consequently, we can control the balance of interactivity and narration by exchanging interactive StoryMap according to the user's context.

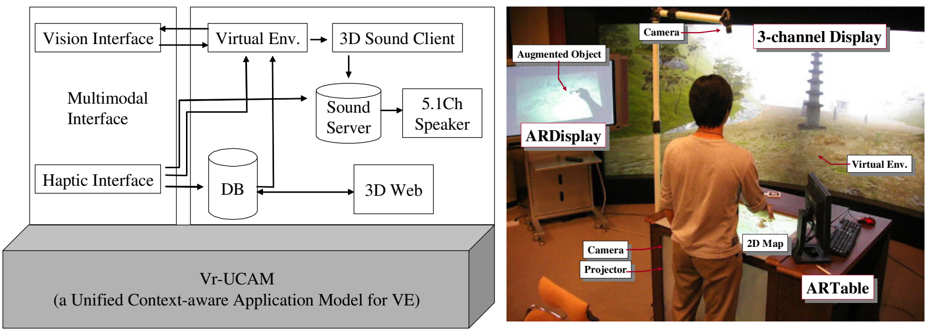
To demonstrate usefulness of the proposed system, we implemented a virtual heritage system which represents legend of *Unju Temple*. Firstly, 2D animation shows legend of *Unju Temple*. Thus, it makes participants understand plot and motif of the historical place. To experience actually the virtual heritage system, users input their profile on PDA, and select a specific doll according to their preference. Then, StoryMap guides the participants to the 7 places which show adaptive events according to the user's profile acquired by PDA. Users can make their own story by manipulating the tangible interface interacting with StoryMap and VE. Accordingly, we believe that the proposed system can be applied to the application areas, such as historical education, interactive entertainment and games, etc.

This paper is organized as follows. In Section 2, we describe detailed components of RMS. In Section 3, we show implementation of the heritage system. In Section 4, we describe a presented story in our implemented system. Finally, the conclusions and future work are presented in Section 5.

2 System Overview of Responsive Multimedia System (RMS)

RMS consists of three key components; a multi-modal tangible user interface (MTUI), a Unified Context-aware Application Model for Virtual Environments (vr-UCAM), and virtual environment manager (VEManager) [7][13]. MTUI allows users to interact with virtual environments through human's senses by exploiting tangible, haptic and vision-based interfaces. vr-UCAM decides suitable reactions based on multi-modal input [6]. Finally, VEManager generates dynamic VE through 3D graphics and 3D sounds, etc.

We integrated MTUI (such as, haptic interface, ARTable, 3D Sound system), VE-Manager, and vr-UCAM into the RMS. Fig. 1(a) shows the structure and describes information flow of RMS. We built Database and Sound server to save 3D Model and to play 5.1 channel sounds. With the haptic interface, we deformed 3D model and delivered it into the DB over the network. We displayed the deformed contents in VE and we downloaded it from the 3D web site. We connected ARTable, Haptic interface and VE with sound server. When triggered signal was delivered to the sound server, it selected proper sound file to play through 5.1 channel speaker. Moreover, we designed 3D web server to allow users to interact our contents over the network. To demonstrate the propose RMS, as shown in Fig. 1(b), we exploited various equipments, such as, three-channel stereoscopic display system, the clustered 3 workstations for cylindrical 3D stereoscopic display, 2 workstations for vision-based and haptic interface, and 1 workstation for sound and database server.



(a) (b)

Fig. 1. Overview of RMS (a) System Architecture and (b) Demonstration of RMS

3 User-Centric Context-Based Storytelling

3.1 User's Context and Its Application

Context is responsible for unfolding the story with sequences of events, execution of events, interaction level, and for changes in VE. Firstly, we can compose different event sequences according to the user's personal information. Secondly, we can offer different events, e.g., if someone selects 'good' mode, he can watch blooming trees. Or others can watch shaking trees by wind. Thirdly, we can also adjust proper interaction level according to users. For example, if a young child experiences a virtual story, we need to restrict degree of freedom of user interface since the child does wring actions as he/she wants. However, if an adult experience a virtual story, we can recommend proper actions for building a story. Finally, we provide users with personalized VE. For example, weather and lightning can be changed according to the user's preference.

The context of a Virtual Storytelling is categorized with explicit context and implicit context. Explicit context indicates a direct command such as gesture, pressing button. Implicit context means personal information or preference. The context supports seamless connection between RE and VE by exploiting as a form of 5W1H [10] [11]. Moreover, we apply two methods to extract context; WPS (Wearable Personal Station) for implicit context, sensors for explicit context. WPS manages the user's needs and preferences [12]. When we want to acquire context by sensors, it depends on target and the sensor type. For example, if we use a vision interface, we can get the user's gesture or position by processing video frames.

3.2 Framework for Context-Based Storytelling

We assume that a virtual storytelling is achieved by interaction between real and virtual environments. From this point of view, we divide the environments into four parts: users, surroundings, virtual objects and virtual surroundings. Story is unfolded with collaboration of users, MTUI (surroundings) in real world and virtual objects, virtual surroundings in VE. As shown in Fig. 2, users, MTUI and VE are connected as an input and output relationship by using vr-UCAM. For instance, multimodal tangible user interface, e.g., dolls and watch, are connected to virtual object (avatars) and time of day in VE. So, when users manipulate user interface, the results are displayed in the VE. On the other hand, an action of virtual object in VE affects states of user interface.

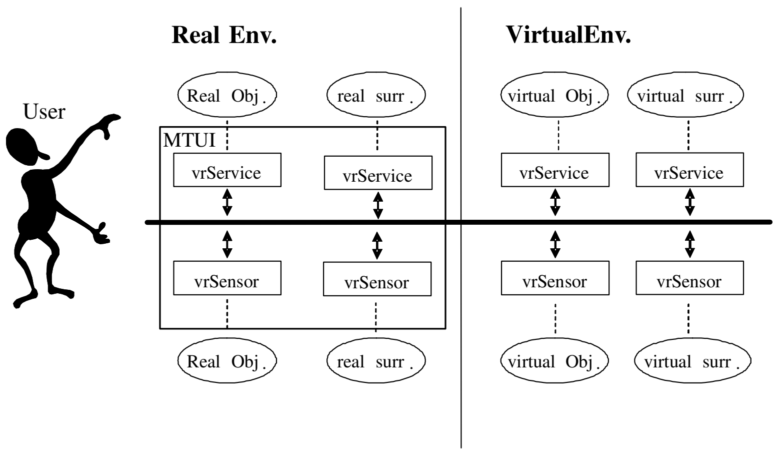


Fig. 2. Relationship between a user, RE and VE. A user interacts with tangible interface in real environment.

vr-UCAM is a framework for designing context-aware applications in virtual environments[6]. It is focused on implementation of intelligent objects which response according to the user's context. To extract the user's context, actual or virtual sensors acquire raw signal, and then generate the preliminary context which only contains limited information according to the sensor's ability. Furthermore, virtual objects and virtual surroundings integrate several contexts from sensors, and determine proper responses. That is, through acquisition, delivery, and processing step, they understand not only explicit input but also implicit input of users. After understanding the user's intention, the applications decide and show pertinent responses.

Based on vr-UCAM, we design architecture for context-based storytelling by exploiting user's profile and preference. As shown in Fig. 2, it consists of users and MTUI in real environments, and virtual objects and virtual surroundings in virtual environments. The users provide personal information to the MTUI and VEManager. Both MTUI and VEManager reflect user's personal information on their application. MTUI guides users to experience a virtual story by composing sequences of events in VE and VEManager controls actions of virtual objects in VE. The author can design MTUI and VE according to story. In this architecture, MTUI, which is composed of Tangible interfaces and Interactive StoryMap, has a role to make a sequence of events in VE by depending on the author's design and the user's interaction.

3.3 Tangible Interface and Interactive StoryMap for Narrative Control

Tangible interface allows users to approach and interact with VE. In details, we classify the tangible interface with media object and control object [8][9]. The media object makes users approach multimedia contents in virtual environments. And the control object offers an intuitive interface which allows users to manipulate VE with their two hands. The tangible interface is designed as a form of objects, e.g., small priest, etc, related to the story.

Interactive StoryMap has an important role to reorganize and adjust a virtual story. The Interactive StoryMap displays images to guide users to author's storyline at the ARTable of RMS [7]. It allows users to interact with VE by manipulating tangible interface and Interactive StoryMap. In VE, there are lots of events at the location. Users read different information on the StoryMap and they visit different ordered locations, since it shows different guide map of VE.

3.4 Action Control of Virtual Object and Virtual Surrounding

Virtual object and virtual surrounding in VE have their own actions which are based on the finite state machine. They are also designed as a form of characters related to the story. For example, let's assume that there are a tree which has three states; standing up (no moving), moving by wind, and blooming flower. We change the states according to the decision of vr-UCAM.

State of a virtual object indicates the action taking by the virtual object. A transition is a condition that the virtual object can turn from the current state to other state. It is expressed $C = < P, T, E >$ to a total model, $P = \{p_1, p_2, \cdots, p_n\}$ is a set of state, and it is expressed as a circle, $T = \{t_1, t_2, \cdots, t_m\}$ is expressed with a quadrangle by a set of a transition. E is a set of arrow and $E \subset (P \times T) \cup (T \times P)$. The state of a virtual object can be shown with m:P ({0,1}. This means that a position and the number of token are put in a circle. Thus, we can express the state of virtual object as $m = [m(p_1), m(p_2), \cdots, m(p_n)]$. For example, let's assume that initial state of virtual object is $m = [1,0,0,0]$. If virtual object receives a triggered signal t_2 from vr-UCAM, the state of a virtual object move to p_2 and it is expressed as $m = [0,1,0,0]$.

4 Implementation

4.1 Legend of Unju Temple

Unju Temple is a temple located in Daecho-ri, Hwasun-gun, Jeonnam Province in Republic of Korea. At the end of Unified Shilla Dynasty, Reverend Priest Doseon tried to construct thousands of Buddhist statues and stone pagodas in a single night. By implementing the history, we try to show wishes of ancient people, who put tremendous amount of effort building Buddhist structures, and the beauty of Unju Temple. We also apply the following scenario: *Users watch animation or listen the legend of Unju Temple. They go to carve a Buddhist statue and station by using haptic interface. After completing to make the Buddhist statue, they move it to virtual environment. They start to navigate virtual Unju temple. There are seven events to unfold the story as user's positions. In addition, users can experience Unju Temple repeatedly through 3D web site.*

4.2 vrSensor and vrService from Story

To design the introduced virtual story, we processed following steps. At first, we assigned events on the seven specific locations. From these events, we designed virtual objects and MTUI which interact with users. Moreover, vrSensor and vrService were implemented and attached to the objects and surroundings. Table 1 shows the context list which can be extracted from implemented vrSensors in the proposed RMS [7].

Table 1. Context List

	Context	Sensor
Real Env.	Coordinate of object in Table	Camera Tracking System
	Time of RE	Time sensor
	Weather of RE	Weather caster or simulator
Virtual Env.	Location of VE	Location sensor in VE
User	Age, gender	PDA
	Preference (animals, color, weather)	User's selection on Table display and tangible objects

Table 2. Seven Events in the implemented virtual Unju

Location	Events description
Great Tree in front gate	When user gets close to Great Tree, the user can hear sounds of wind and birds.
Road of Pagoda	Priest Doseon appears and disappears repetitively to guide direction
Cherry blossom	Cherry reacts according to user's gesture and context.
Field of Small Pagoda	User can station small pagoda and wish his/her desire. It records voice of the user.
Construction Rock	Stars twinkle up in the sky. Pagoda and Buddhist twinkle on the ground.
Maeyeorae	User can watch ruins of Maaeyeorae
Lying Buddhist	Priest Doseon appears and then user can watch lying Buddhist at one sight from higher position. When enough number of pagodas is stationed, lying Buddhist starts to stand up.

In addition, Table 2 shows events based on the legend of Unju temple. We designed user's interaction focusing on 'field of small pagoda' and 'lying Buddhist' since we want to show the desire of ancient people. We also categorized the RMS into real surrounding, virtual object, virtual surrounding, regarding the scenario shown in Table 3. Then, we described interaction between these four elements according to the user's context.

Table 3. Actions of vrSensors and vrServices

vr-UCAM	Name		Actions
vrSensor	Real surrounding	vrController	Senses a user's gesture from motion of object
		vrWeatherforecaster	Detects current weather state
		vrTimer	Acquires current time
	Virtual surrounding	vrLocationSensor	Senses location of a user's avatar in virtual Unju
vrService	Real surrounding	vrViewer	Displays maps or pictures on MTUI
	Virtual object	vrNavagator	Changes the camera path
		vrDoseon	Appears or disappears when he want.
		vrMaae	Begins to Ruin according to time.
		vrPlant	Grows and Blooms
	Virtual surrounding	vrWeather	Changes the color of sky, location of sun, fog in VE
		vrSound	Generates specific sounds

4.3 Tangible Object, Interactive StoryMap, and Events in VE

We implemented Multi-modal Tangible User Interface to unfold virtual story. We made several kinds of dolls, e.g., a small priest, a tiger, and a dragon, etc, related to the story, and participants can select one of them according to their preferences. For building stone pagodas, we made shovel-shape object. So, participants can move stone with shovel-shape object. Furthermore, we connected a physical watch with time in VE, and then we used the watch to allow participants to change position of sun in VE.

We also implemented interactive StoryMap which was displayed on ARTable of RMS as shown in Fig. 3. It was designed for guiding or causing user's interaction. The StoryMap showed geometric information of 'Unju Temple'. When participants invoked interaction, specific position lighted up to guide next position to go. It can show different story since a author or a developer can change sequence of event.

We implemented Virtual Plant, Priest Doseon, and ruining Maaeyeorae. We created key frames of virtual object and combined them as a sequence by using pfSequence. A pfSequence is a pfGroup that sequences through a range of its children, drawing each child for a specified duration. Each child in a sequence was thought of as a frame in an animation. We controlled whether an entire sequence repeats from start to end, repeats from end to start, or terminates.

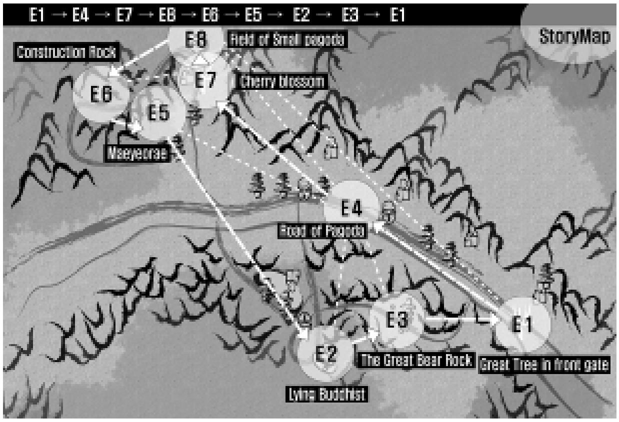


Fig. 3. Interactive StoryMap on ARTable

4.4 Scenario of Virtual Storytelling Based on User's Profile and Preference

At first, participants inputted personal information on PDA instead of WPS. We used this personal information to initialize overall system. i.e., we determined sequence of location, events, interaction level, and surrounding of VE. For example, if the participant was a young child, we selected a simple sequence of event and low interaction level because the young child was difficult to follow full story. And the child's interaction was restricted specific position which guide presented story. If the participant was an adult, we showed recommendation of next position, and then the participant could select the next event. Participant's preferred color and mood affect the weather in VE. If the participant chose one of rainbow color, one of mood (bad, good), it changed the color of sky and fog, raining, sun shining in VE.

The participant started interaction with objects and interactive StoryMap on ARTable. The participant watched objects, which is easily used in our daily life, on the table. There were three kinds of dolls (small priest, tiger, and dragon) for navigation and a bottle for pouring water. The participant could select one of dolls. If he selected the tiger, he crept toward front as in a view of tiger in VE. If the small priest was selected, screen view slightly moved up and down as human walking. If the dragon was selected, screen view rendered at the view point of flying dragon.

The first position where a participant put his/her doll was the entrance of 'Unju Temple' at StoryMap. When the participant arrived at the entrance, the entire map was changed to the detailed map. In the map, the participant can navigate near the entrance as he/she wants. The participant watches big tree moving by wind, and listen a sound of bird moving around the tree. When the participant placed the doll at specific position on the map, he exited out to entire map. Then, he recognized the recommended position with shining effect. He can select other position for further navigation.

After watching scenery on the path, the participant moved to the location where cherry blossom was planted. As explained above, the state of cherry was initialized according to the user's profile. We implemented three state; steady, shaking, flower. At this position, the participant poured the water to the plant by using object on table. Then the cherry bloomed beautiful flowers.

The next location was the field of small pagodas. It represented the desire of ancient people. At this location, the participant piled up small stones in VE and recorded their personal desire during 10 second. The participant grabbed an object which has a shape of shovel and holds a small stone on the table display. After the participant built a small pagoda with three stone, recoding voice started. The recorded voice file was saved to DB server. Finally, the small stone pagoda was moved to the field of stone pagodas in VE.

And the participant experienced several events at the next location. When the participant arrived at the rock of construction, he can watch the whole sight view of 'Unju Temple' and special event occurs. The pagodas and Buddhist status on the road twinkled at the road and it showed constellation. At Maaeyeorae, the participant watched the ruinous statue. It tells us absence of a thousand years.

Finally, the participant reached at lying Buddhist. According to the legend, when thousands of Buddhist statues were made, the lying Buddhist stands up right. In our story, when 10 small stone pagodas are made by participants, the lying Buddhist stands up right and recorded sound at the field of small pagoda played. After ending, the participant can visit any place and any events in VE.

5 Conclusions and Future Work

In this paper, we presented the context-based storytelling with Responsive Multimedia System (RMS). In our proposed approach, users can interact with VE intuitively through implemented tangible interfaces. Moreover, users can experience a personalized story in which events, interaction level and scene of VE are reorganized and adjusted according to a user's profile and preference. Consequently, we can control balance of interactivity and narration by changing interactive StoryMap according to the user's profile. In the future, we have plans to find evaluation method to improve our storytelling approach. In addition, we need to develop authorized tools to make contents and stories with RMS. Furthermore, we will apply the concept of artificial life for the virtual object.

References

1. M. Mateas, "An Oz-Centric Review of Interactive Drama and Believable Agents," *AI Today: Recent Trends and Developments*, M.Wooldridge and M.Veloso, eds., LNAI, vol. 1600, pp. 297-328, Springer Verlag, 1999.
2. R.M. Young, "Notes on the Use of Plan Structures in the Creation of Interactive Plot," *Amer. Assoc. Artificial Intelligence fall Symp. Narrative Intelligence*, AAAI Press, Menlo Park, Calif., 1999.
3. N.M. Sgouros, G.Papakonstantinou, and P.Tsanakas,"A Framework for Plot Control in Interactive Story System," *Proc. Am.Assoc.Artificial Intelligence*, AAAI Press, pp. 162-167, 1996.
4. M. Cavazza, F.Charles, and S.J.Mead,"AI-Based Animation for Interactive Storytelling," *Proc. IEEE Computer Animation*, IEEE Press, pp.113-121, 2001.
5. M. Cavazza, F. Charles, and S.J. Mead, "Character-based Interactive Storytelling," *IEEE Intelligent Systems, special issue on AI in Interactive Entertainment*, pp. 17-24. 2002

6. Sejin Oh, Youngho Lee, Woontack Woo, “vr-UCAM2.0: A Unified Context-aware Application Model for Virtual Environments,” *ubiCNS*, 2005.
7. Youngho Lee, Sejin Oh, Beom-chan Lee, Jeung-Chul Park, Youngmin Park, Yoo Rhee Oh, Seokhee Lee, Han Oh, Jeha Ryu, Kwan H. Lee, Hong Kook Kim, Yong-Gu Lee, JongWon Kim, Yo-Sung Ho and Woontack Woo, “Responsive Multimedia System for Context-based Storytelling,” *Proc. of Pacific-Rim Conference on Multimedia*, 2005.
8. Sejin Oh, Woontack Woo, “Manipulating multimedia contents with Tangible Media Control System,” *ICEC*, Vol.3166, pp. 57-67, 2004.
9. H. Ishii, B. Ullmer, “Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms,” *in Proc. of ACM CHI*, pp. 234-241, Mar. 1997.
10. Dey, A.K, and Abowd, G. D. The Context Toolkit: Aiding the Development of Context-Enabled Applications, *In proceedings of the 1999 Conference on Human Factors in Computing System*, 434-441, 1999.
11. S. Jang, W.Woo, “Unified Context Representing User-Centric Context: Who, Where, When, What, How and Why,” *ubiComp workshop (ubiPCMM)*, 2005. (accepted)
12. Y. Suh, W.Woo, “Context-based User Profile Management for Personalized Services,” *ubiComp workshop (ubiPCMM)*, 2005. (accepted)
13. Dongpyo Hong, Woontack Woo, “I^2-NEXT: Digital Heritage Expo,” *ICAT04*, pp. 120-125, 2004.

FELIX 3D Display: Human-Machine Interface for Interactive Real Three-Dimensional Imaging

Knut Langhans, Klaas Oltmann, Sebastian Reil, Lars Goldberg, and Hannes Hatecke

Youth Research Center, Vincent-Luebeck-High-School,
Glueckstaedter Strasse 4, 21682 Stade, Germany
mail@felix3d.com
http://www.felix3d.com

Abstract. Flat 2D screens cannot display complex 3D structures without the usage of different slices of the 3D model. A volumetric display, like the FELIX 3D Display can solve this problem. It provides space-filling images and is characterized by "multi-viewer" and "all-round view" capabilities without requiring cumbersome goggles.

The FELIX 3D Displays of the swept volume type use laser-light to project real three-dimensional images upwards a rotating screen. Because of some disadvantages using rotating parts in this setup, the FELIX Team started investigations also in the area of static volume displays. The so called, "SolidFELIX" prototypes, have transparent crystals as a projection volume. The image is created by two or one IR-laser beams.

The projected images within all FELIX 3D Displays provide a fascinating, aesthetic impression through their inherent, unique three-dimensional appearance. These features of a 3D display could be combined in an interface between a virtual reality scene and the real world. Real-time interactions and animations are possible. Furthermore, the display could host an intelligent autonomous avatar that might appear within the display volume.

Potential applications as a virtual reality interface include the fields of entertainment, education, art, museum exhibitions, etc.

The FELIX 3D project team has evolved from a scientific working group of students and teachers at a normal high school in northern Germany. Despite minor funding resources considerable results have been achieved in the past.

1 Introduction

Up to today researchers are seeking for ways to display our three-dimensional world in an appropriate way: In all three physical dimensions, interactively, in real time, and in true color. Two-dimensional display systems are still not able to realize depth cues like perspective or hidden contours satisfyingly.

For an ideal dynamic 3D image representation a technique should be preferred with a multiple user, all-round view capability, without the usage of visual aids. It should also satisfy all depth cues such as stereo vision and motion parallax. Volumetric displays, which generate 3D images within a physical volume rather than on a stationary surface, meet most of these needs. The images are thus placed within the

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 22 – 31, 2005.

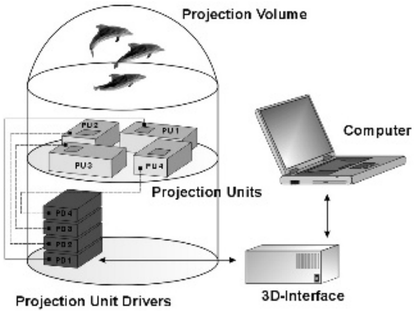


Fig. 1. Modular concept of FELIX 3D display (swept volume)

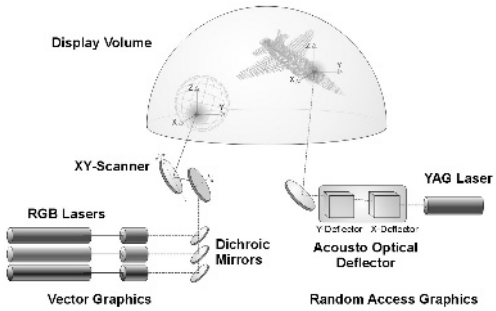


Fig. 2. Different projection techniques for swept volume displays

physical world of the observer, in comparison to virtual reality systems where the observer is placed within the virtual environment of the non-physical image space. The development of all our FELIX 3D Displays is characterized by its comparatively simple system configuration. One or a combination of the three projection techniques can be implemented: vector graphics, raster graphics, random access graphics (Fig. 2).

Moreover all FELIX displays are compact, light, and easy to transport systems of modular design and consist of standard components (Fig. 1).

2 Basic Concepts of Volumetric Imaging

To produce volumetric images, it is necessary to create a display volume of some type within which the image can be drawn. A number of attempts have been made in past and current development efforts. We generally divide these attempts in two basic categories:

- Swept volume displays,
- Static volume displays.

Fig. 7 gives an overview of the basic classes with examples, which do not claim to be complete. Some approaches of related developments will be described in more detail in the following sections. Holographic techniques are not discussed in this paper as they represent a separate class of its own.

2.1 Swept Volume 3D Displays

In the case of swept volume displays, volumetric images can be created within a display volume by reflecting or transmitting light from a moving screen, which is oscillating or rotating. A periodically time-varying two-dimensional (2D) image is used to sweep out the volume cyclically at a frequency higher than the eye can resolve. Eventually a spatial image is formed through persistence of vision. Thus, several of the usual cues of everyday depth perception are exhibited, especially the compelling ones of stereoopsis and parallax.

The primary 2D pattern may be generated on either an emissive panel or a passive projection screen. The depth effect is achieved when the screen surface, which can take various shapes, is oscillated perpendicularly to itself, or when it is rotated, in synchronism with the cyclic 2D pattern. Alternatively, the screen may be fixed but viewed in an oscillating plane mirror, which has a similar effect.

2.2 Static Volume 3D Displays

Systems which are able to create a display volume without the need to employ mechanical motion are classified as static volume displays. The goal of this technique is to provide emissive dots of visible light (so called *Vo*lume Pi*xels* or voxels) at a large number of locations in a static setup incorporating no oscillating or rotating parts. Several interesting attempts have been made using transparent crystals, gases, electronic field sequential techniques and others.

The phenomenon of stepwise excitation of fluorescence seems to be one suitable and promising approach for the generation of isolated fluorescent voxels.There are two different principles of excitation: Two frequencies, two step (TFTS) upconversion and one frequency, two step (OFTS) upconversion. The TFTS method uses two laser beams of different wavelengths, which intersect within the display material, whereas the OFTS method uses one laser beam of a constant wavelength, which has to be strongly focused to produce a voxel.

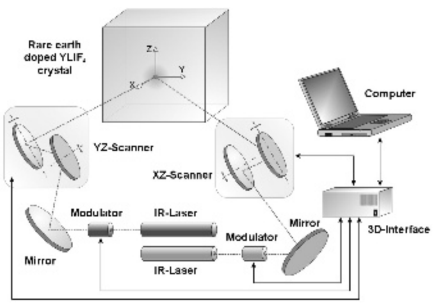


Fig. 3. System Architecture of the SOLID-FELIX 3D Display (static volume; TFTS)

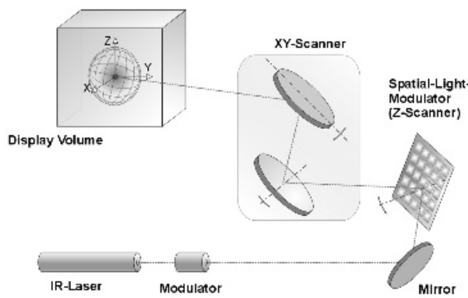


Fig. 4. Principle of a static volume OFTS display, using a spatial-light-modulator[23]

Both techniques are based on the absorption of two photons (normally invisible IR photons) by optically active atoms, molecules or ions. Translucent materials doped with certain ions or phosphors can emit voxels, when exited by two or one infrared laser beams.

In the following sections we will illustrate the past and present developments in the field of volumetric imaging by selecting representative examples for different approaches.

3 Historical Overview

3.1 Swept Volume 3D Displays

Swept volume systems have been discussed theoretically since 1918. However, one of the first functional prototypes of volumetric displays was presented by the early pioneers Parker and Wallis in 1948 using a CRT to create simple 2-D pattern which were projected onto a vibrating mirror or a rotating flat screen. More elaborated techniques, presented by the ITT Laboratories in 1960, consist of an especially programmed high brightness cathode ray tube (CRT) whose blips are optically transferred to a translucent rotating screen within a glass cylinder (Fig. 5). Later, Robert Batchko described a similar system where he used a vector-scanned laser illumination source instead of a CRT. Inspired by these research activities, Gregg Favalora (Actuality Systems, Inc.) recently set up a volumetric display, implementing state-of-the-art hard- and software. A high resolution projection engine illuminates a diffuse projection screen that rotates with projection optics and corresponding relay mirrors.

In another arrangement, introduced by R. Ketchpel in 1962, a phosphor-coated flat screen rotates in a vacuum, with a controlled electron beam striking its surface (Fig. 6). Based on this technology B. Blundell, A. Schwarz et. al. from the University of Canterbury (Christchurch/New Zealand) presented a volumetric display, called "Cathode Ray Sphere CRS", in 1992.

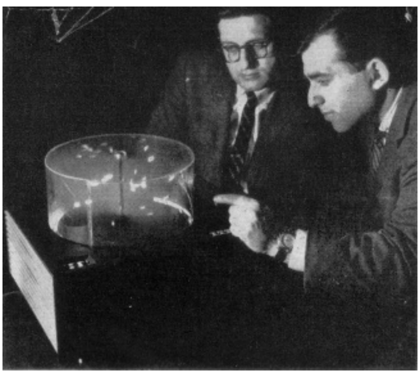


Fig. 5. CRT Projection on rotating screen (ITT, 1960)[17]

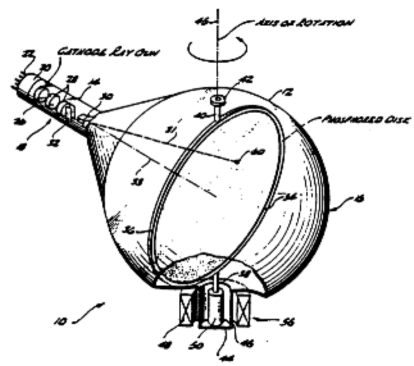


Fig. 6. Rotating phosphor-coated disk in CRT (Ketchpel, 1962)

In 2002 Jang Doo Lee et. al.[4] from Samsung, Korea, presented a paper about “Volumetric Three-Dimensional Display Using Projection CRT”. Within the display they used a flat target screen.

An early approach using a curved moving surface was described in 1962 by D. W. Perkins. He introduced the "spherical spiral display" with a special shaped screen that rotates about its vertical axis and onto which a light beam is projected from below by an optical system. The intention was to use it as a 3D radar display. In 1965 A. M. Skellet filed a patent application with a rotating tilt upward screen.

As a representative of the category of swept volume techniques, we describe the general principle of operation used in our 3D display FELIX. In 1983 we started investigations of different prototypes of the "FELIX 3D Display". These systems refer to the concept of projecting laser images on a rotating screen. Our investigations in this field mainly focus on different screen shapes (helical, double helical, flat), a modular and portable setup, advanced projection techniques and software development.

Volumetric 3D Displays

Swept Volume Techniques

Flat Surface

- high brightness CRT projection (Parker / Wallis, 1948; ITT Lab. 1960)
- "Generescope" (M. Hirsch, 1958)
- electroluminescent panel (R. J. Schipper, 1961)
- Cathode Ray Tube Imager (R. D. Ketchpel, 1962)
- LED array (E. Berlin, 1979)
- Cathode Ray Sphere (B. G. Blundell et. al.,1992)
- laser projection (R. Batchko, 1992)
- 3D Rotatron (S. Shimada, 1993)
- rotating reflector (C. Tsao, 1995)
- System and Method For High Resolution Volume Display Using a Planar Array (Douglas W. Anderson, et.al 1995)[9]
- Actuality Systems (G. Favalora, 1996)[16]
- CRT (Jang Doo Lee, et. al., 2002)[4]

Curved Surface

- "Spherical Spiral Display" (D. Perkins, 1962)
- “Cathode ray tube for three-dimensional presentations” (A.M. Skellett, 1962)[7]
- helical mirror (W.D.Chase, 1976)
- **helical shaped screen**:
  - de Montebello, 1969
  - Szilard,J., ultra sonics, 1974[6]
  - R. Hartwig, 1976
  - "FELIX 3D-Display", 1983
  - R.D. Williams, 1988 "OmniView"
  - R. Morton, 1990
  - P. Soltan, 1992
  - NEOS Technologies, 1995
  - Zheng J Geng, 2001[8]
  - Live 3-D Video in Volumetric Display (Jung-Young Son, et. al., 2002)[5]
- **Archimedes’ Spiral**:
  - de Montebello, 1969 "Synthalizer"
  - H. Yamada, 1986

Oscillating Screen

- vibrating CRT (~ 1940)
- phosphor screen in CRT (E. Withey, 1958)
- moving mirror (~ 1960)
- reciprocating screen (C. Tsao, 1995)

Static Volume Techniques

Solid

- **TFTS upconversion**:
  - fluorescence in $CaF_2:Er^{3+}$ with use of filtered xenon lamps as excitation sources (J. Lewis et.al., 1971)
  - Stereoscopic projection technique (E. Luzy and C. Dupuis, 1912)
  - Display system using rare-earth doped fluoride crystals (M. R. Brown, G. Waters, 1964)
  - Display system based on simultaneous absorption of two photons (M.A. Dugay, et. al., 1970)
  - Three-color, solid-state, three-dimensional display (E. Downing, et. al., 1994)
  - SOLID FELIX 3D Display, 2000[14]
  - Principle demonstration of 3D volumetric display achieved by PrYb co-doped materials (Xiaobo Chen, et. al., 2000)[13]
  - Volumetric display based on up-conversion phosphors, (Jianying Cao, et. al., 2005)[22]
- **OFTS upconversion**:
  - One-beam pumping of Er^{3+}-doped ZBLAN glasses (T. Honda, 1998)
  - SOLID FELIX 3D Display, 2000[14]
  - Dispersed crystallite up-conversion displays (M. Bass, et.al. 2001)[12]
  - One-beam pumping upconversion volumetric-display (X. Chen, 2002)[19]

Gaseous

- Intersecting laser beams in Rubidium vapor (E. Korevaar, 1989)
- Electron beams intersecting in nitrogen gas filled image space (F. S. Howell, 1945)
- Multiple-step excitation in mercury vapor (R. Zito, Jr., 1962)
- Two-step excitation of fluorescence in iodine monochloride vapor (R. H. Barnes, et. al., 1974)
- Imaging phosphor display using voxel activation by two crossing laser beams (E. Downing, B. Torres, 1989)

Fig. 7. Overview of volumetric imaging techniques (for detailed references see [1,14])

In the FELIX approach the primary 2D image is generated on a passive projection screen. The rotating screen sweeps out a cylindrical volume. This volume provides a three-dimensional display medium in which scanned laser pulses are projected. The hitting laser beam will be scattered from the rotating surface causing a voxel. The spatial position of the emanating voxel within the display is determined by the momentary location of the laser beam's intersection with the rotating screen.

The 2D image is generated by one or more projection units from below the display volume. A projection unit consists of a light source and a XY scanning system. As a light source we use a monochromatic laser as well as a laser-based red-green-blue (RGB) color mixer. Depending on the application we use different types of scanning systems (galvanometric, acousto-optic and micro-electronic-mechanical). Thus, this setup makes it a powerful and flexible tool to keep track with the rapid technological progress of today.[1,2,14,23]

In 1988 R. D. Williams and F. Garcia demonstrated a display wherein a scanned laser beam is displayed upon a continuously rotating disc that is mounted on a motor shaft at an angle. Later they also employed a helical surface in their device further known as the "OmniView 3D Display".

Since early 1990 a research team (P. Soltan, M. Lasher et. al.) at NRaD, the RDT&E Division of the Naval Command, Control and Ocean Surveillance Center (NCCOSC, San Diego), in cooperation with NEOS Technologies, works on a similar development as well using a helical surface.

In 2001 Zheng J. Geng[8] filed a patent “Method and Apparatus for an interactive volumetric Three-Dimensional Display” providing a process and system for interactively displaying large (more than one million voxels) volumetric 3D images utilizing a sequence of helical slices of a 3D data set. Core elements of his setup are a spatial light modulator and a helical target screen.

3.2 Static Volume 3D Displays

The effect of TFTS upconversion has been known since the 1920s, when it was first observed in mercury vapor, later it was used for 3D laser display approaches.

Based on the pioneering work of J. D. Lewis et. al. of Batelle Laboratories in 1971, E. Downing et. al. at Stanford University, presented in 1994 a three-color 3D display with improved upconversion materials using high power infrared laser diodes.

Another approach, patented by E. J. Korevaar in 1989, used a gaseous volume enclosed in a sealed glass container.

In 1998 Honda investigated the quantum efficiency for OFTS upconversion in 0.2 mol% Er^{3+} doped ZBLAN glass. This alternative excitation method uses one focused laser beam at a wavelength of 979 nm.

Since 1999 the FELIX team is also investigating static volume approaches.Core element of our setup is the display volume, which consists of a cubic transparent material (glass or crystal), doped with ions from the rare earth-group of the elements. In our case the primary dopant is Er^{3+} in a concentration of 0.5 mol%. These ions are excited in two steps by one (OFTS) or two (TFTS) infrared laser beams. Initial tests with our experimental setup, called "Solid FELIX", revealed promising results.[23]

In 2001 M. Bass[12] from the University of Central Florida filed a patent for a OFTS static 3D volume display in which a dye doped display medium is used, where the

Miscellaneous	
– stylus in gelatine material (Chrysler, 1960) – 3D fiber optic display (A. Gery, 1983) – voxel addressed by optical fiber (D. L. Macfarlane, 1994) – field-sequential projection on stacked electrically switchable mirrors (T. Buzak, 1985) – scattering LC-display (Q. Luo, et. al.) – Intersecting electron beams in a phosphor particle cloud (W. G. Rowe, 1975) – 3D optical memory (D. A. Parthenopoulos, P. M. Rentzepis, 1990)	– layered LC-based 3D display using one light source (M. S. Leung, et.al., 1998) – layered LC-based 3D display with several light sources (R. S. Gold, J. E. Freeman, 1998) – Multi-planar volumetric display system using psychological vision cues (Sullivan, et. al.,2001)[11] – Development of 3-D display system by a Fan-like Array of Projection Optics (Toshio Honda, et. al., 2002)10 – 3D Display using parallax barrier: “The SeeLinder” (T. Yendo, 2004)

Fig. 8. Miscellaneous volumetric display techniques[1]

emitting particles are dispersed in a transparent host (acrylic plastic or glass). He proposed a spatial light modulator to scan the focus fast enough through all three dimensions of the display cube.

In the following years extensive and elaborated research work on static volume 3 D displays was done by Xiaobo Chen, et. al.,[19,20,21] from Beijing Normal University and Institute of Physics (China) concerning appropriate materials as well as different excitation methods (OFTS and TFTS),scanning methods and multiple color mixing.

Apart from the described technologies several other interesting attempts have been made in the field of static volume displays. Early approaches from the beginning of the 1960s proposed the mechanical insertion, support, and translation of objects or particles in a volume.

An alternative static volume procedure, suggested by D. Macfarlane in 1994, proposed to use a three-dimensional stack of volume elements, made of an UV-cured optical resin doped with an organic dye. Each voxel is addressed by an optical fiber that pipes light to the voxel.

4 Interface Techniques

A basic problem of interactive virtual reality is the interface between the virtual 3D world and the real 3D world. All current known standard interfaces are only 2D interfaces, like a flat monitor or the computer mouse. Reducing 3D to 2D always means losing information. An approach of solving this problem is the usage of 3D interfaces.

Volumetric 3D displays will take the first barrier, the reduction of the third dimension of the displayed scene. The communication and interaction with the virtual reality will also be increased by using such a display. Projecting a real 3D image boosts the number of interaction methods and devices. Many solutions and approaches were presented in the past, but mainly as theoretical ideas. Only a few 3D interfaces came on the market. Most of the launched interfaces are focused in the area of CAD.

Fig. 9. Dolphins swimming through the display

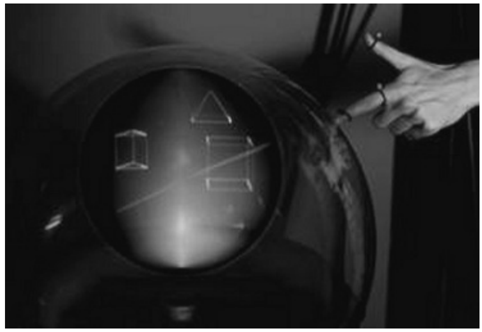


Fig. 10. Multi-finger gestural interaction device using visual tracking methods[18]

A direct interaction with the volumetric display should not require any additional interface devices, like a mouse, a keyboard or a additional flat monitor. In our opinion, visual tracking techniques are the most desirable approaches, because of their advantages of intuitive interaction. For instance, the visual tracking of hand gestures is qualified for 3D interactions, due to the fact that people use additional gestures to express things, feelings and shapes, they are talking about.

The FELIX group started investigations in the field of input devices and interaction possibilities. As an input device for volumetric displays we discussed a six degrees of freedom device, operating in three dimensions. The device is mechanically connected to the display's mounting and connected to the user's arm-wrists. Sensors track the motion and forces of the users hand or arm and translate it to commands for the display.[1] To improve the interactive use of volumetric displays we also worked with avatars, using artificial intelligence. The vision is to create a three dimensional head in the display, which acts and behaves like a human being.[15]

Balakrishnan et. al.[3] summarized the features of different user interface approaches and discussed the interface design by using physical mockups. They recommend "separated spaces" (display volume and interface volume are separated) for interface approaches and also discuss nonhaptic interaction, like eye-tracker and voice controlling.

As a conclusion it can be said that an interface for a volumetric display should be highly interactive and must consider that people got used to interact with a real 3D world in their normal life. A 3D Interface should adapt as many as possible of these interacting techniques to offer an intuitive handling of the controlled virtual 3D world.

5 Areas of Application

Spatial visualization of data can simplify and accelerate the working process. As described above, volumetric displays can be used as an interface for interactive virtual 3D scenes. With an adequate 3D user input device, following applications are possible:

- Scientific visualization: *Projection of multi-dimensional data structures*
- Entertainment: *Demonstration of FELIX as an interactive 3D play-console*

– Computer Aided Design: *Collision analysis for virtual mockups*
– Chemistry/physics/pharmacy: *Visualization of molecular structures*
– Medical Imaging: *Display of 3D ultrasonic pictures*
– Air Traffic Control: *3D visualization of a typical air traffic scenario*
– Interface for virtual reality: *highly interactive interface device*

Fig. 11. Animated models in the Felix 3D display

Fig. 12. Possible scenario: Discussion with a 3D avatar in the volumetric display

Acknowledgements

The authors wish to thank Prof. Ernst Heumann and Dr. Klaus Petermann (University of Hamburg), Prof. Andreas Tünnermann (Fraunhofer IOF and University Jena), Prof. Gerd Litfin (Linos AG), Dr. Ulrich Peuchert (Schott AG), Prof. Holger Zellmer (University Jena, now HTWK Leipzig), Prof. Dietmar Müller and Ralf Seidel (Chemnitz University of Technology), Hermann Neumann (University Stuttgart), Y. Oldenburg (GSI Lumonics), Thomas Engel (Linos-Phontonics GmbH), Jan Brune (LISA), Dr. Wilfried Stuettgen (Sun Microsystems GmbH), Tovi Grossman (University of Toronto), Elisabeth Rieper and Christian Guill (Youth Research Center). This work was supported by the Klosterkammer Hannover.

References

1. K. Langhans, D. Bahr, D. Bezecny, D. Homann, K. Oltmann, K. Oltmann, C. Guill, E. Rieper, G. Ardey, "FELIX 3D Display: An Interactive Tool for Volumetric Imaging", Proceedings of SPIE, Vol. 4660, San Jose, CA, 2002.
2. K. Langhans, D. Bezecny, D. Homann, D. Bahr, C. Vogt, C. Blohm, and K.-H. Scharschmidt, "New portable FELX 3D display", Projection Displays IV, Proceedings of SPIE, Vol. 3296, San Jose, CA, 1998.
3. R. Balakrishnan, G. Fitzmaurice, G. Kurtenbach, "User Interfaces for Volumetric Displays", IEEE Computer, March 2001
4. Jang Doo Lee, et. al. "Volumetric Three-Dimensional Display Using Projection CRT", Proceedings of SPIE Vol. 4660, 2002

5. Jung-Young Son, et. al. “Live 3-D Video in Volumetric Display”, Proceedings of SPIE Vol. 4660, 2002
6. Szilard,J., “An improved three-dimensional display system”, ultra sonics, 273-276,1974
7. A.M. Skellett, “Cathode ray tube for three-dimensional presentations”, US Patent 3204238 filed July 13, 1962
8. Zheng J Geng, “Method and Appartus for an Interactive Volumetric Three Dimensional Display”, US Patent 6900779, filed Jan. 26, 2001
9. D. W. Anderson, et. al., ”System and Method For High Resolution Volume Display Using a Planar Array”, US Patent 6208318, filed Jan. 12, 1995
10. Toshio Honda, et. al., “Development of 3-D display system by a Fan-like Array of Projection Optics”, Proceedings of SPIE Vol. 4660, 2002
11. Sullivan, et. al., “Multi-planar volumetric display system and method of operation using psychological vision cues”, US Patent 6466185, filed Aug. 20, 2001
12. M. Bass, et. al., “Dispersed crystallite up-conversion displays”, US Patent 6654161, filed July 31, 2001
13. Xiaobo Chen, et. al., “Principle demonstration of 3D volumetric display achieved by PrYb co-doped material”, Proceedings of SPIE Vol. 4221, 2000
14. K. Langhans, C. Guill, E. Rieper, K. Oltmann, D. Bahr, “SolidFELIX: A Static Volume 3D-Laser Display”, Stereoscopic Displays and Applications XIV, Proceedings of SPIE, Vol. 5006, Santa Clara, CA, 2003
15. Felix 3D Display homepage, http://www.felix3d.com
16. G. Favalora, “Volumetric 3D Displays and Application Infrastructure”, IEEE Computer, August 2005
17. N.N., “3-D display”, Space/ Aeronautics, pp. 60-67, Sept. 1962
18. T. Grossman, D. Wigdor, R. Balakrishnan, “Multi-finger gestural interaction with 3D volumetric displays”, UIST 2004 Proceedings, ACM Symposium on User Interface Software and Technology, p. 61-70, 2004.
19. Xiaobo Chen, et. al.,“Physics basis of one-beam pumping upconversion three-dimensional volumetric display based on erbium-doped glass”, Cockpit Displays IX: Displays for Defense Applications, Proceedings of SPIE ,Vol. 4712, pp. 426-435,2002
20. Chen, Xiao-Bo, et. al., “Up-conversion luminescence research of Er(0.5):ZBLAN material for volumetric display application when excited by 1520 nm laser”, Chinese Physics, Volume 12, Issue 12, pp. 1451-1458, 2003
21. Chen Xiao-Bo, et al, “Excited-State Upconversion of Pr(0.5)Yb(1.5):ZBLAN under Two-Colour Excitation”, Chinese Phys. Lett. 22, p.963-966, 2005
22. Jianying Cao, et. al., “Design and implementation of an advanced three-dimensional volumetric display based on up-conversion phosphors, Cockpit and Future Displays for Defense and Security; Proc. SPIE Vol. 5801, p. 278-286, 2005
23. H. Hatecke, K. Langhans, Youth Research Center ,Vincent-Luebeck-High School, Stade, Photonik 5/05, pp.2-5,2005

Proposing Daily Visual Feedback as an Aide to Reach Personal Goals

Ana C. Andrés del Valle and Agata Opalach

Accenture Technology Labs, 449 Route des Cretes - BP 99 - 06902,
Sophia Antipolis, France
ana.c.andresdelvalle@accenture.com

Abstract. Personal goals are difficult to achieve and healthy habits are not as simple to maintain as people would like to. Often, humans look for help to change unwanted behavior. Technology could help us to overcome the daily struggle faced during the search for behavior improvement. The Persuasive Mirror is an example of how computerize persuasion, in the form of visual feedback, could give the needed support.

1 Introduction

Mrs. Novak knows that regular exercise and a balanced diet is good for her both in the short and in the longer term. However, in her day to day life she often forgets to include a walk and usually prepares the same quick and not always nutritionally balanced meals. She wishes she had a personal coach who would help her vary the recipes while keeping them simple (she no longer has the patience to prepare elaborated dishes) and to help her stick to regular exercise. She buys a Persuasive Mirror and has it configured to match her nutritional and fitness goals.

One morning, when Mrs. Novak brushes her teeth, her Persuasive Mirror shows a healthy looking face and displays a reminder that she has a doctor's appointment at 10am. It also confirms that the weather forecast is good and suggests she takes a stroll to the doctor's office. In the evening, Mrs. Novak finds that her face looks more tired than usual. With a wave of a hand she sees detailed statistics on the behavior that day: that's right, she drove to the clinic and had a frozen pizza for dinner, ignoring the system's suggestion for a salad with ham and pine nuts... A year later, Mrs. Novak is very happy with her increased energy, surely due to the regular exercise she now has and her balanced diet. She half-laughingly maintains that her healthy lifestyle is "to keep her mirror happy" but she is also proud to announce to her daughter that her doctor confirmed osteoporosis improvement so she can take up dance classes again at her senior club.

Mrs. Novak's story represents one of the scenarios where a Persuasive Mirror [2], Accenture Technology Labs concept mirror that gives personalized feedback to

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 32–40, 2005.

the user in order to persuade him to pursue certain goals, could be beneficial. Very often, people undergo psychological struggle when willing to keep up a healthy and reasonable lifestyle. Often, the activities that "*please our minds*" do not match those that are "*healthy for our body*". Changing unwanted behavior has become a concern. For instance, in 1998, one out of ten books sold in the US was related to self-help –*The Wall Street journal, December 8, 1998.* It seems, therefore, that psychological support might be needed to help people that at some moment in time are determined to change their habits because they want to improve their quality of life.

The Persuasive Mirror is a fine example of computerized persuasion. Computerized persuasion or *captology*, is the study of computers as persuasive technologies [9]. It explores the design and applications of interactive products created for the purpose of changing people attitudes or behaviors in some way. The *captology* efficiency of a piece of technology is strongly related to the capability that its interface has to impact the user. Ideally, the selected technology has to be non-intrusive, often used and able to provide the user with relevant feedback linked to the activities he is concerned about. We considered several physical forms to give shape to the concept, and finally decided to use a mirror because it is a natural object in people's life and it is generally used more than once a day. The Persuasive Mirror provides daily motivation aide to individuals because it is a "nagging" object that uses sensors to gather information about our behavior, analyzes these data with relation to our goals, and visually rewards good and exposes bad behavior.

Next Section contains an overview of the Persuasive Mirror architecture. We go over the technical challenges of replicating a mirror using digital reflection in Section 3. In Section 4, we review several useful visual feedback that could be used. Sections 5 and 6 cover social aspects related to the persuading nature of the mirror. First, we discuss some of the psychological strategies relevant to our work; second, we expose how the system needs to be customized to every user's characteristics and goals. We conclude our article by discussing some future work.

2 Overview of the Persuasive Mirror Architecture

The Persuasive Mirror is an augmented mirror that we have designed with three working modes: 1) progress representation, showing feedback on recent behavior (e.g. silhouette slimmer and younger for a balanced meal, bigger and older for no meal or junk food), 2) behavior summary (statistics on "good" and "bad" behavior, areas to improve), and 3) a regular mirror (equivalent to switching the device off).

Fig. 1 depicts the architecture behind the device. Three modules integrate the system:

Sensor network: Sensors are the backbone of any intelligent technology. The sensor network plays a critical role in the architecture because it gathers information about human activity in various ways. The mirror reflection is designed based upon the user's behavior. Activity related to health habits can be observed, for instance, using video cameras [1], that study global activity; UV cameras that analyze sun damage of the skin; vital signal analyzers; smart fridges that control the amount of food that is taken from them; or, smart balances, that not only measure mass but also give out the percentage of fat contained in our bodies.

Control system: The control system can be considered the brain of the Persuasive Mirror. It acts as a bridge between what the person does and what the mirror provides as a feedback. Upon receiving the data from the sensor network it must decide the kind of visual feedback the digital mirror will give to the user. As we explain in Sections 5 and 6, there is no universal strategy for persuasion that can be applied to everyone. In fact, the control of the mirror is the result of complex customization of the global system based on the experts advice about health and psychological influence. The mirror is a user-centered designed device, therefore, the control system will require the feedback of the user's evolution and reaction from the moment the device is utilized.

Digital mirror: Altered visual feedback of our own face can have great impact on us. In order to be able to give customized reflection a flexible digital mirror has been created. The digital mirror not only works as a regular mirror but also enhances the user's face thanks to different Image Processing techniques.

The complete architecture allows the Persuasive Mirror to act accordingly to the user's goals and behavior.

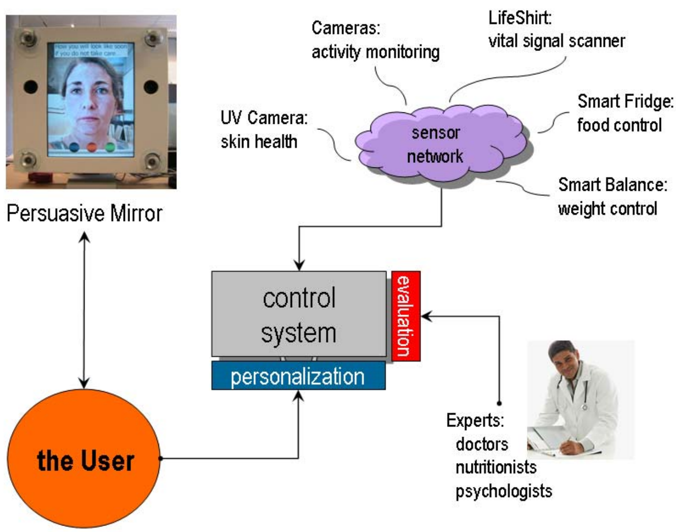


Fig. 1. The Persuasive Mirror architecture

3 Digital Replication of a Mirror

The first technical challenge faced was to create a system that replicates a mirror. This challenge involves several geometrical and computational aspects that have to be solved. When somebody is looking at a mirror, he can see the scene from different viewpoints, depending on his position. To replicate the reflection, this viewpoint can be obtained by rendering the image from a virtual camera whose position is symmetric to the user over the mirror axe.

Taking into consideration these geometrical aspects, some researchers at the University of Southern California [10] developed one approach, using a tracker and a single camera. The tracker returns the relative position of the face. Then the system knows what part of the scene given by the camera has to be displayed on the screen. Nevertheless, as read in the previous articles, some regions that have to be displayed might be occluded to the camera (Fig. 2). Therefore, to avoid occlusions, we should use more than one camera. In our particular case, to use the simplest settings, we have decided to use two cameras situated on the left and right sides of a screen. Our prototype is pictured in Fig. 3.

Starting from this configuration, we will have to solve two main problems. First we need to track the eyes of the user. A real time technique has been developed by Darrell et al. [7]. Applying usual tracking techniques on the output of two cameras, the 3D position of the face is computed by triangulation. It is also possible, as shown in [15], to not only find the eyes position but also the gaze direction.

The second issue is to derive a single image that describes the scene from the right viewpoint using two cameras. One way to achieve this goal is described in [4]. Several cameras are used to display one defined target in the scene. For each camera, a list of penalties due to several parameters, like the distance to the target or the resolution of the camera, is defined. Then, in the final image, every pixel is coming from the corresponding best camera. A second approach is described by Criminisi et al. [6]. The developers use only two cameras, and apply a modified Dynamic Programming (DP) algorithm to compare the images. Using defined cost functions, the DP algorithm computes for each line of the 2

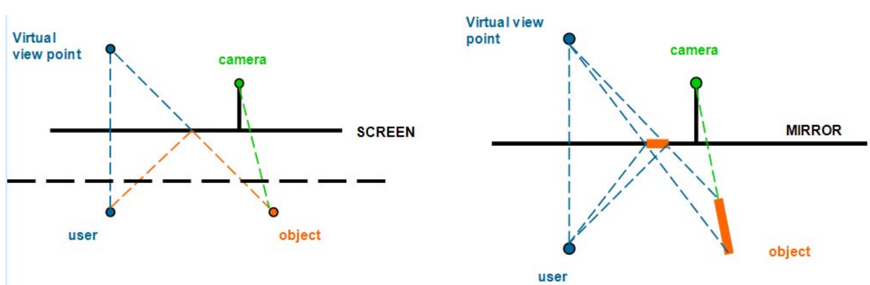


Fig. 2. These diagrams show how the mirror reflection can be simulated (left) and the limitations of using a single fixed camera because of the occlusions

Fig. 3. Components of the digital mirror: two cameras, one touch screen and some lights

images the best alignment called the minimum cost path. Every pixel of this path is projected on the 45-degree line called the ”virtual scanline” using the position of the virtual camera, defined for replication as a center of projection.

Finally we can see in [3] a large study of the most used algorithms in stereo computing. Local methods like ”Block Matching” or ”Feature Matching” compute a pixel by pixel matching. These methods are very fast but several artifacts appear due to occlusions. Global Methods like ”Graph Cuts” or ”Dynamic Programming” compute global matching. The results are better but these methods are more computationally expensive.

After evaluating the current state of the art, we decided to extend the work of [6] to mix our 2 inputs. The DP algorithm seems to give acceptable results and we have specially adapted it for real-time applications. We will also use triangulation in order to compute the 3D position of the eyes that will define the location of the virtual camera.

4 Advanced Image Processing to Generate Feedback Enhancements

Image enhancements are essential to provide visual feedback. They can go from intelligently exposing data, to altering the user face appearance according to the effect we want to produce and the message we want to convey. For instance, morphing techniques like Face Transformer [8] can be used to render the face younger or older. Subtle reminders of the consequences of the user's behavior can also be conveyed by artificially playing with colors, non-photorealistic rendering or augmented reality techniques.

Other possibilities are to use augmented reality (mixed reality) techniques in order to add artificial elements into the scene. An additional difficulty in this case is to match the lighting on the artificial elements and the lighting of the real scene in order to make the additions believable.

Experimentation is needed to evaluate which enhancements are the most effective for any particular user. One general outcome of not doing exercise and failing to have a healthy lifestyle is the gain of weight. As a first enhancement, we are developing image processing algorithms based on Active Appearance Models, facial-anatomical deformation constrains [17] and skin color changes to show the user the consequences of unhealthy habits on his reflection before they happen (e.g. fattening).

Fig. 4. A time line example of the reflection the user will get with the Persuasive Mirror. In the morning (a-b), suggestions will be given according to the user's personal goals. At night, the mirror will review the user's activities regarding the proposed suggestions; it will give visual feedback to let the user understand the consequences of her acts.

5 Psychological Strategies to Achieve Persuasion

In order for the mirror to perform its persuasive role, the visual feedback given will have to match the psychological strategy aimed at creating the wanted reaction in the user. The goal of designing a persuasion strategy is motivating the mirror's owner to have a healthier lifestyle. Let us stress that the mirror has no pre-encoded universal values it tries to impose on the user. It is the user that configures the device in order to receive daily assistance in achieving their goals.

Some of the psychological concepts we would like to take into account are [5]:

- **Punishment versus reward:** the mirror reflection is able to present the future consequences of a certain behavior. It will show the bad results of

not behaving appropriately versus the good results of acting correctly. The objective being that the user will be motivated by seeing that a change in behavior really matters.
- **Likes and dislikes:** people have more positive or negative preferences to certain colors, shapes, and visual structures. The mirror could take advantage of that by relating the visual environment the user's likes and dislikes. Likes will be attached to positive behavior; dislikes to negative.
- **Positive reinforcement:** [16] is the technique used to encourage people to start having or maintain a certain behavior by reminding them how good it is to them (positive feedback) but also how it can still be improved (by letting them know that the final goal is getting closer). Once a certain positive behavior has appeared, it is challenging to maintain the interest of the user in continuing acting appropriately. A punishment versus reward strategy has proved not to be always optimal because continuous reminders about future negative consequences might discourage people. That is why the all strategies have to be integrated inside a positive reinforcement framework.

Fig. 4 presents an example of how the mirror could change the perception of our activities by showing us the consequences of our acts. In this case, the mirror has two different attitudes: (a) a controlled diet has allowed the user to lose those kilos she did not want, and the mirror reminds her the progress she has made (*positive reinforcement*); and (b) the user has been under the sun without protection and has not followed the mirror's advice, the mirror shows her the future consequences of her unhealthy act (*punishment* & *dislikes*).

6 Personalization to the User's Needs

Most processing steps behind the mirror will have to be adapted to each individual using it because what motivates people is a very personal matter. Given a set of psychological strategies and a certain number of persuading visual-enhancement techniques, the Persuasive Mirror system has to be tuned to apply those that will effectively persuade the user. No individual is and reacts alike, the adaptation and customization of the Persuasive Mirror to the user's profile is a major issue.

The technical challenge in the customization field is to build a system that automatically configures to the user requirements with little (ideally none) user interaction. As a starting point in our research, we propose to proceed incrementally, offering very slight comments first and exaggerating them on request from the user until the optimal configuration is achieved. Eventually, previous knowledge about the user's taste and psychology could help to speed up this customization stage.

7 Conclusions and Future Work

The Persuasive Mirror was conceived as part of the Accenture's Intelligent Home Services initiative [12]. This initiative aims at enabling the elderly to age autonomously at home. To allow older people stay healthy, specific coaching can be given in the form of visual feedback. In an environment where artificial intelligence and sensors will be deployed all over the home [13], [14], [11], this mirror would become a personal coach.

The Persuasive Mirror is a visually enhanced display that alters people's reflection to motivate reactions and change in behavior. We believe that computerized persuasion delivered by such a device can help the users reach their personal goals. Interface design is central in captology devices. It becomes even more crucial if the interaction between the device and the person aims at producing a change in the user's behavior.

Any computerized system provokes the same initial fear: "Can we trust a computer?" This concern becomes even stronger when the device performs critical tasks. The Persuasive Mirror could play a major role in the user's life. The owner of the mirror will have to be reassured that its operation is correct and that, above all, it is completely fault tolerant. This fault-tolerance is an important part of the system as the effectiveness of the device will be determined by the correct processing of the input data. Beyond the natural fears associated with technology, the use of persuasive devices raises ethical concerns regarding their potential to drive people to act in a certain manner. As it happens with any technology, its uses and abuses must be evaluated. Like self-help books, we believe that if it is used with explicit consent, the outcomes can be beneficial.

Researchers at the Accenture Technology Labs have been mainly focused on the technical challenges. Our future research lines will include how to create the automatic personalization of the mirror feedback. Automatic customization of technology can be of great interest to different industries.

We are also interested in finding what psychological strategies are the most suitable. We want to establish collaborations with experts in the medical and psychology areas. They could help us perform field studies that could shape the strategies behind the Persuasive Mirror.

References

1. Activity monitoring. Intelligent Home Services. Accenture Technology Information retrieved on August 2005 from: http://www.accenture.com/xd/xd.asp?it=enweb&xd=services\technology \research\ihsactivity.xml
2. Andrés del Valle, A. C., Opalach, A.: The Persuasive Mirror: computerized persusion for healthy living. Human Computer Interaction International (2005) LV, US.
3. Brown, M. Z., Burschka, D., Hager G. D.: Advance in Computational Stereo. Department of Computer Science, the Johns Hopkins University.
4. Buehler, C., Bosse, M., McMillan, L.: Unstructured Lumigraph Rendering. Proceedings of SIGGRAPH (2001)
5. Cialdini, R. B.: Influence. Science and Practice. Allyn & Bacon Publishers, MA. (2001) ISBN 0-321-01147-3
6. Criminisi, A., Shotton, J., Blake, A., Torr, P.: Efficient dense stereo and novel-view synthesis for gaze manipulation in one-to-one teleconferencing. [Technical Report] MSR-TR-2003-59, Microsoft Research (2004)

7. Darrell,T., Gordon, G., Woodfill, J., Harville, M.: A Virtual Mirror Interface using Real-time Robust Face Tracking. Proceedings of the the Third International Conference on Face and Gesture Recognition, IEEE Computer Society Press, (1998) Nara, Japan.
8. Face Transformer: Morphing-based face transformer at Perception Laboratory, School of Psychology, University of St Andrews, Scotland. Retrieved March 29, 2005 from: http://www.dcs.st-and.ac.uk/ morph/Transformer/index.html
9. Fogg, B. J. (2003). Persuasive technology: using computers to change what we think and do. Morgan Kaufmann Publishers
10. François, A., Kang, E., Malesci, U.: A Handheld Virtual Mirror. Integrated Media Systems Center, University of Southern California.
11. Hutchinson, H. et al. Technology probes: inspiring design for and with families. Proc. ACM CHI 2003 Conference on Human factors in Computing Systems (2003) 17–23
12. "Intelligent Home Services" – Accenture: Accenture Technology Labs Intelligent Home Services initiative. Retrieved March 25, 2005 from: http://ww.accenture.com/ihs
13. Intille, S.: Designing a home of the future. IEEE Pervasive Computing (2002) 80–86
14. Intille, S., Larson K.: Designing and evaluating supportive technology for homes. Proc. IEEE/ASME International Conference on Advanced Intelligent Mechatronics
15. Nakanishi, Y., Fujii, T., Kiatjima, K., Koike, H.: Vision-Based Face Tracking System for Large Displays. Proceedings of the Ubicutous Computer International Conference (2002) Göteborg, Sweden.
16. Pryor, K.: Don't shoot the dog! The new art of teaching and training. [**Chapter 4**] Bantam Books (1999) ISBN 0-553-38039-7
17. Terzopoulos, D., Waters, K: Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models. IEEE Trans. Pattern Anal. Mach. Intell. 15(6) (1993) 569–579

Sound Navigation in PHASE Installation: Producing Music as Performing a Game Using Haptic Feedback

Roland Cahen, Xavier Rodet, and Jean-Philippe Lambert

Ircam 1, Place Igor Stravinsky, 75004 Paris, France
+33 1 44 78 48 68
{Roland.Cahen, Xavier.Rodet,
Jean-Philippe.Lambert}@ircam.fr

Abstract. Sound Navigation consists in browsing through different sound objects and sound generators situated within a virtual world including virtual spatialized sound and visual scenes, to perform a musical trajectory and composition. In the PHASE Project installation, the 3D virtual world resembles the surface of a vinyl disk, magnified so that one can see the groove and move a "needle" (the "reading head") in it and out of it to read the disk. Another such "needle" (the "writing head") can "write" music in the groove. A part of the game is a pursuit between the writing head and the reading head handled by the player. Different musical devices have been implemented. Most of them have a haptic behavior. The scenario is fully related to the musical metaphor and aims to give equivalent pleasure to contemplative as well as to competitive players.

1 Context of the Project

PHASE (Plateforme Haptique d'Application Sonore pour l'Eveil musical) is a research project devoted to the study and the implementation of multi-modal systems for generation, handling and control of sound and music. Supported by the RIAM network, it was carried out by the CEA-LIST and Haption (for the haptic research and supply of the haptic device), Ondim (for the development of visual content and system integration(and Ircam (for the research and developments about sound, music and the metaphors of interaction). These developments were implemented in an interactive installation for the general public in the form of a musical game integrating various metaphors. CNAM-Cedric has been involved in the game scenario and its ergonomics. During the last three months of the project, a demonstration of some of the results of the PHASE project in the form of an interactive installation offering the public a musical game was presented and evaluated at the G. Pompidou museum in Paris. The aim of the game is to play music and to develop musical awareness. The pedagogical aspect was thus significant for children and for adults. This installation was an extraordinary success; many players of all age and different cultural backgrounds appreciated it very much, including blind people, which shows the validity of such a device and opens the way to many new possibilities of gestural music control.

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 41 – 50, 2005.

1.1 Project Purposes

The objectives of the PHASE project are to carry out interactive systems where haptic, sound and visual modalities cooperate. The integration of these three interaction channels offer completely innovative interaction capacities, comprising gesture with force and tactile feedback, 3D visualization, and sound spatialization.

The objectives are scientific, cultural and educational:

– To study the cognitive processes implied in musical exploration through the perception of sound, gesture precision, instinctive gestures and the visual modality,
– To try out new ways of interactivity (haptic interaction, spatialized sound and 3D visualization, in real time),
– To propose new sorts of musical gestures, and by extension, to enrich the methods of training of technical gestures in any field (industry, handicraft, education, etc.),
– To improve the characterization of sound and its relationship to music (exploration of the multidimensional space of timbre) and to offer a new kind of listening,
– To implement new methods of sound generation and to evaluate them with respect to a public of amateurs and professionals,
– To propose new forms of musical awareness for a large audience, specialized or not.

Finally, an additional objective was to test such a system, including its haptic arm, in real conditions for the general public and over a long duration in order to study and to measure and improve its robustness and reliability; and its interest for users.

2 Haptic Interface

2.1 Design Objectives

The haptic interface used in the project had to allow natural and intuitive manipulation and exploration of sound and music. Therefore, it must respect following design objectives:

– Performance criteria: The user must feel as if s/he is acting directly in the virtual environment. S/He must feel free in an unencumbered space (which requires a large and free workspace, low inertia and low friction) and feel crisp contacts against obstacles (which requires llarge force feedback, bandwidth and stiffness).
– Integration criteria: The input device must be compact enough to be easily integrated in a workstation or exhibition area.
– Maintainability criteria: It must be designed with commercially available components as far as possible in order to minimize cost and simplify maintenance. Consequently, a Virtuose 3D 15-25 haptic interface has been deigned by French society Haption [14]. It is thus well adapted to our specific needs. It offers a rugged and validated basis well adapted to the PHASE project specific needs. Since the installation was conceived to be used by the general public over a long duration, particular care was paid to its safety, its robustness

and its reliability. Moreover, to allow for an elbow-supported manipulation as previously specified, an ergonomic desk was designed and manufactured at CEA- LIST. It enables a precise and comfortable use without fatigue. It also allows an immediate understanding of the workspace of the arm, thus making the handling intuitive.

3 Gesture, Music and Metaphors

In the Phase project, gesture is the entrance point of the considered multi-modal systems. In view of gesture interaction, playing a traditional instrument is the first idea, which comes when thinking of playing music, but there are many others ways such as controlling music, for example conducting or mixing. A series of studies were thus carried out to measure the relevance and the properties of some gestures in the specific context of musical control.

Fig. 1. Modified Virtuose 3D 15-25 integration

More generally, we searched which freedom could be left to the player to express himself/herself in a given musical space, in order to produce music at a satisfactory expressive level. One can, for example, have access to rate, rhythm, timbre, tonality, etc. Various modes of gestural and musical playing have thus been studied. The aim for the user is to play and control music. The user's gesture is carried out in the real space (containing his hand). Various feedbacks (haptic, sound and visual) influence his gesture. For example, to help the user to play a musical instrument, which requires long practice before precise control is reached, haptic feedback can be used as a guide like sliding on a string. Note that a scenario of the course as in the installation's game can also facilitate training. Unlike vision and hearing, the haptic feedback loop is direct since gesture and force feedback are localized in the hand of the user. Among other important necessities, coherence of the various modalities must also be guaranteed between the real and virtual world. This is achieved by using a metaphor which defines the link between the real world where the hand acts, and the virtual music world, where the avatar acts. According to the context, one can define various metaphors, which lead to different implementations, more or less close to reality. Several types of metaphors were tested. Some of them will be detailed in section 5.

3.1 Playing Music

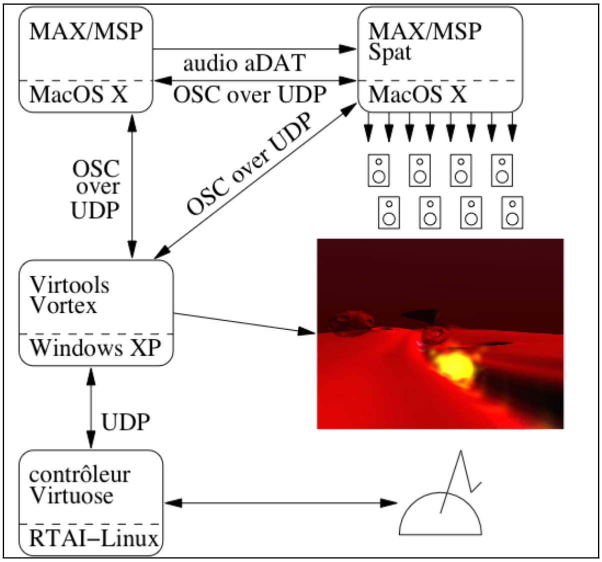


Fig. 2. Architecture

One of the innovative aspects of the PHASE project is the aim of generating, handling and controlling sound and music. Different ways of using gesture for playing music have been studied. On one side, the player can, for example, have access to rhythm, timbre, tonality, etc., and on the other side different gestures can be used for this control. Different modes of playing (from a gestural and musical point of view) have been identified: positioning oneself in the music, browsing the music, conducting the music, playing the music (like with an instrument, etc.). There are also different kinds of music that can be played and re-played. We use the term replay meaning that the music is already recorded or programmed, to be precise, the player does not exactly compose it nor interpret it, as he or she is did not study the score before. Replaying strategies can be either very simple and passive or very active and creative. As explained in the previous section, well-known music and temporal musical structures can easily be identified even when they are quite affected by a transformation, and the control of a temporal position is easy. However, music is otherwise fixed and not adjustable in such a scheme. On the contrary, specific music pieces can be composed so as to be modified by players. Furthermore, music processes or generators can be designed and programmed in order to produce a specific kind of music in order to interact with players in the particular way needed for the application.

Different musical interactive tools have thus been realized, that allow users to:

_ Manipulate musical phrases, specifically created and easy to transform.

_ Change musical parameters in play time or/and real time, for creating different variations.

_ Generate a continuous staccato flow, omnipresent and typify of an object or a character, enabling numerous programmed and interactive variations very near to real time. Some work on recorded or generated sound material has also been realized in the project using, for instance, granular synthesis which offers a privileged access to time control. It allows the system to re-play the music in different ways and to separate pitch transposition from time position and speed control. For example, one can control directly the temporal development with gestures. One can also control a tempo, or acceleration and deceleration. Manipulating short time musical structures modifying time window in loops or on a stream. Another approach consists in constructing a virtual physical world and listening to it in different ways. It is then tempting to exploit the measures that come from the physic engine directly, particularly with the so-called physical model synthesis. Finally, the musical link between different physical and sound objects needs to be defined and developed. Writing and playing such sounds and music, using dynamic controls remains a challenging task.

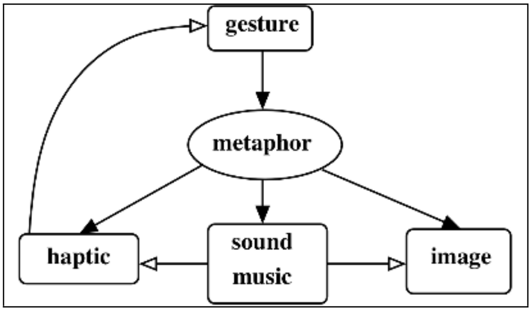


Fig. 3. Interaction metaphor

3.2 Sound Navigation

The trajectory and interaction of the user in a simulated world articulates music events and phrases. This is called *Sound Navigation*. *Sound Navigation* consists in browsing through different sound sources placed in a specialized sound scene to compose a musical form from the mix produced by movements and interaction with these sources. It is an artistic transposition of our sound experience in the physical world.

The user or player can freely choose between a simple visit, play with sound behaviours or perform a personal expressive path.

The choice of a musical implementation of each sound actor related to its interaction metaphor, is determined by the musical functions attributed in the composition. As the player can choose to trigger different sound behaviours in its own choice order, the horizontal construction is determined by the environment interaction setup. For example, two sounding objects in different places cannot be hit at the same time, and there is a speed limit between hitting the two objects successively depending on the distance and the possible velocity of the controller. Sound navigation applies to musical situations where the spatial metaphor offers non-conventional musical paradigms. These situations have a strong spatial organisation and therefore should not be confused with image illustration, for they are not images and do not illustrate anything other than themselves.

3.3 Gesture and Haptic Feedback

On account of the complexity of the whole system, a recording and mapping abstract module has been realized in order to disconect gesture control from various implementations, allowing to switch from a controller to another. This allows testing musical or sound behaviours and interactive gestures without the haptic arm itself, which necessitates technically complex realizations. For example, a graphic tablet or a simple mouse has been used for quicker experiments. One obviously needs the complete device for final adjustments, but the abstract module permits to validate concepts and implementations under simulation. Naturally, the absence of force-feedback, the changes of the controller body and it's set of parameters must be taken into account in such a simulation.

In addition, it is possible to record gestures, to re-play them (in order to test, regulate and to validate dynamic sound behaviours) or analyze them offline, with or without the haptic arm. Thanks to the assistance of Frederic Bevilacqua, a study was carried out to compare a graphic tablet (Wacom Intuos A4 and aerographer) and a force-feedback arm (Virtuose 6D from Haption, www.haption.com). The goal was to identify significant parameters in the context of real-time sound and music generation. In addition, this study was used to evaluate the limitations of the graphic tablet against the haptic arm within the context of our metaphors.

4 Scenario

While in the large majority of cases, the music of video games is put at the service of visual, a part of our research was devoted to the use of the three modalities for the processes of musical awareness: how to design a device to sensitize general public with new listening and musical creativity modes?

To answer this question, a simple scenario has been conceived with a very intuitive beginning [16], [17]. It is based on the metaphor of a writing head and a playing head in a vinyl disc groove. This metaphor thus involves three main actors, the two heads and the human player, and allows for a significant amount of possibilities of interaction.

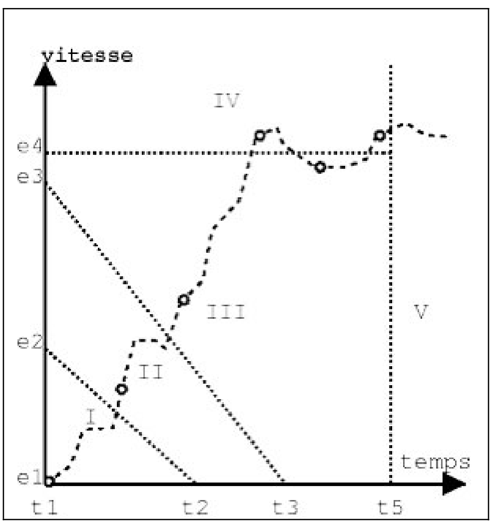


Fig. 4. Gameplay schematics

Liliana Vega and Thomas Gaudy did the game design itself, under the direct-ion of Stephan Natkin from CNAM-Cedric. One of the main issues of a game shown to any public in a museum, including musician and not, children and adults, experienced players and art amateurs, was to allow different kind of playing. We considered gamers as competitive players and others as contemplative ones. As evaluating music production quality is quite a tough achievement for a computer, we decided to evaluate the player using two simple parameters: time and speed. Every player will go through level I to III and at the end of the time to level V but only excellent player will reach level IV. Game scenario and music.

Level 1 : The machine (the writing head WH) traces a curve corresponding to a musical phrase. The player must reproduce it holding his avatar (the reading head RH). If he succeeds, a new phrase curve is produced more difficult and progressively linking becomes continuous.

Level 2 : The imitation becomes a chase and the ground starts rolling. The player can also wander around experimenting sound and musical interactions.

Level 3 : Once the player has reached the writing head the scenery changes, the landscape becomes more accidental and more obstacles and traps come out to slow down the player. Each of these obstacles produce sound when interacting.

Level 4 : When the player’s avatar reaches the writing head, they stay magnetized and move together in a duo, until too much opposite force tears off their link. Only good players can reach this level and are trying to stay in it as long as they possible because it is very fast and the music becomes virtuoso and dramatic.

Level : 5 is the end. All forces disappear, everything slips until the end/start position of the game is reached again.

Unfortunately only levels II and III where implemented and it was not possible to implement levels I, IV and V and theses parts of the scenario require a future opportunity to be realized.

The musical aspect of the metaphors corresponding to the scenario will be detailed in section 5.

In PHASE, music and player actions are strongly attached together through the gameplay. In other words the score is the game itself or its rules and the interpretation comes out of the player's involvement in the action. The musical result is then deeply dependent on the scenario and on the physical/visual world.

The mode of competitive play is involving for the player but is not imposed. At any time, it is possible to stop the RH, leave the groove and to play with sounding objects or zones of interaction, which are always sounding, even during the race. This helps immersion and sound and musical awareness for the player, by the richness of the sonorities and the proposed interaction. When the player stops, the WH waits for him/her in silence and starts again as soon as the player comes back into the groove. There is no real accurate "scoring" of the player's performance, but the progression of the player is evaluated in order to adjust the behaviour of the system: a player having difficulties to play and a contemplative one are different, but in both cases the system slows down, thus facilitating the usage of the system for everyone.

4.1 Scenario Implementation

One of the main clues was to be able to test the gameplay and to adjust the scenario while testing. In this aim, Xavier Boissarie has created a Virtools scenario editor, which allowed us to assemble groove bricks and build sequences and to precisely position the obstacles. We wished we could dispose of a meta tool for elaborating, defining, modelling and testing such interactive scenario, instead of trying to link bits of systems together. Liliana Vega (CNAM Cedric) has made some pre-figuration of such tool. At the end, different parts of the scenario are divided between each functional elements of the system.

5 Musical Creation in the Installation

5.1 Metaphors

In the PHASE installation, we have mainly developed 3 interaction metaphors: the obstacles, the zones and the music groove. The three of them can play all together or separately. We decided that they should have a good sound behaviour independently and a complementary orchestral function when they play together. Moreover, their sound behaviour has not only esthetical functions but also informative functions, such as saying whether the player's avatar is touching the groove's bottom, or the ground, informing the player that a new obstacle is coming soon when visually hidden, indicating excessive pressure on an object, etc.

5.2 Obstacles: Direct Interaction

Direct interaction resembles ‘sonification’. The physical model is taken as a starting point to produce a significant immersion for the player, since the modalities are then close to those usually perceived in reality. In this case, physical objects are also sound sources. The interaction between the user and these objects is similar to the play of a musical instrument.

Obstacles are objects in the game space with various shapes. They are separated into two types: fixed and mobile obstacles. They have several sound behaviours: Approach, hit, press, scrub and bounce.

5.3 Zones : Divertimento Variations

Zones are musical playing spaces on the sides and inside the curves of the groove. They are represented by long grass moving when touched. They allow a musical real time control on an animated dynamic sound process according to the position inside the zone, the speed of input in the zone, the pressure and the velocity of movement.

5.4 Music Groove: Main Clue

The writing head (WH), represented by a bright spot of light, runs in the groove at a certain distance behind the player. It is flashing synchronously with music. One hears and sees the writing head (WH), which generates and plays music according to the player's actions and the position in the scenario.

The writing head plays the music and writes it into the groove in the form of a visual and haptic trace made of colour traces corresponding to notes.

The player holds its avatar, the reading head (RH), which is represented by another spot of light pursuing the writing head (WH) when placed inside the groove. The reading head replays the music after a delay which is a function of the distance between the writing and the reading heads. It flashes synchronously with the music events read from the trace. The RH pursues the WH.

The trace links the RH and the WH together. The WH writes the trace in the groove, which scrolls when the RH is inside it, like a needle in a vinyl disc groove. Thus, the movement of the RH is relative to the trace and the player controls indirectly the speed of the re-played music. To follow or even catch up the WH, the player must follow the music as exactly as possible along the trace.

The choice of this metaphor is justified for several reasons:

_ An analogy with the real world: one can think of a vinyl disc the groove of which is engraved by the WH and re-played by a needle held in one's hand (the latter situation does not happen in the real world)
_ A well-known game mode: a race between to vehicles (here the WH and the RH ; an improbable situation)
_ A musical principle of re-play, close to the canon or the counterpoint (here quite a strange canon)
_ Facility of gesture : the groove serves as a guide for the player.

6 Music and Public Reception

While visitors using the installation in free access at the G. Pompidou museum were playing music, or skimming through R. Cahen's composition, their performance was recorded on a CD that was given to them at the end of their session (5 to 20 minutes). The installation was an extraordinary success. The number of visitors is estimated at approximately 20.000, including children, adults, musicians, composers, teachers, etc. The reaction of the public was constantly enthusiastic, underlining the interest, the innovation and the pedagogical aspect. The realization of this interactive installation for the general public in the form of a musical game integrating various metaphors shows the value of such a device.

7 Conclusion

The PHASE project was dedicated to research and experiments on the use of gesture control with haptic, visual and sound feedback, with the aim of playing music. Hardware, software and methodological tools were designed, in order to allow for the development and the realization of metaphors having a musical purpose. These developments are now available and usable. They were integrated in a installation used in an open public area with a great success.

One essential characteristic of the installation is the scenario. The metaphor of the vinyl disk has proved to be easily understandable by the public. Manipulating the reading head by mean of the haptic arm is very intuitive. The pursuit is a familiar situation which is related to a musical process. All these properties allow the user to be both in a game situation and playing music. This is one of the main results of the research work, since keeping the interest of the users and gamers to a musical creation activity was one of the most difficult challenges at the start of the project.

In the scenario, the gameplay is nearly transparent in the sense that the user does not need to be conscious of the different levels he/she is going through. It does not require any preliminary training.

Haptic game design reveals itself to be able to contribute significantly to music expression. Considering an haptic play area as an imaginative navigation score seems to considerably enhance the game experience.

Such research projects, linking scientists, engineers and artists, as the ones the RIAM promotes, appear at early states of development, to be a good way to improve technology and new forms of expression.

It opens the way to many possibilities of gesture music control, original fields in music composition, interpretation and game design. A DVD has been realized (in French) which presents the project and the installation.

References

[1] A. Hunt, M. Wanderley, R. Kirk, "Toward a Model for Instrumental Mapping in Expert Musical Interaction", Proceedings of the International Computer Music Conference (ICMC), Beijing, China, October 1999

[2] R. Cahen, "Générativité et interactivité en musique et en art électroacoustique", 2000, http://perso.wanadoo.fr/roland.cahen/Textes/CoursSON.html/MusiqueInteractive.htm

[3] M. Wanderley, "Interaction Musicien-Instrument : application au contrôle gestuel de la synthèse sonore", Ph.D. diss. (in French), University of Paris-6, 2001.

[4] C. Cadoz, M. M. Wanderley. Gesture - Music. In M.M. Wanderley and M. Battier, editors, Trends in Gestural Control of Music, Ircam - Centre Pompidou, 2000, pp. 71-94.

[5] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of the movement", Journal of Experimental Psychology, 47, 1954, pp. 381–391.

[6] J. Accot, S. Zhai, "Beyond fitts' law: Models for trajectory-based hci tasks", In ACM/SIGCHI: Conference on Human Factors in Computing Systems (CHI), 1997.

[7] T.H. Massie, J.K. Salisbury, "The PHANToM haptic interface: a device for probing virtual objects", Proceedings of the ASME Winter Annual Meeting, Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, Chicago, November 1994.

[8] K. Young Woo, B.D. Jin, D.S. Kwon, "A 6-DOF force reflecting hand controller using the fivebar parallel mechanism", Proceedings of the 1998 IEEE International Conference on Robotics and Automation, Louvain, Belgium, May 1998, pp. 1597-1602.

[9] R. Baumann, R. Clavel, "Haptic interface for virtual reality based minimally invasive surgery simulation", Proceedings of the 1998 IEEE International Conference on Robotics and Automation, Louvain, Belgium, May 1998, pp. 381-386.

[10] D.A. McAffee, P. Fiorini, "Hand Controller Design Requirements and Performance Issues inTelerobotics", ICAR 91, Robots in Unstructured Environments, Pisa, Italy, June 1991, pp. 186-192.

[11] G.W. Köhler, "Typenbuch der Manipulatoren - Manipulator Type Book", Thiemig Taschenbücher, Verlag Karl Thiemig, München, 1981.

[12] G. Burdea, P. Coiffet, "La réalité virtuelle" (in French), Hermès Publishing, Paris, 1993.

[13] F. Gosselin, "Développement d'outils d'aide à la conception d'organes de commande pour la téléopération à retour d'effort", Ph.D. diss. (in French), University of Poitiers, June 2000.

Virtual Characters

Action Planning for Virtual Human Performances

Markus Löckelt

DFKI GmbH, Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany
loeckelt@dfki.de
http://www.dfki.de/~loeckelt

Abstract. We describe an action planning approach that integrates a system of autonomous dialogue engines with narration and presentation modules to create a multi-modal storytelling environment. The goal is to let multiple virtual and human participants act on equal footing in a story that allows interactions that are flexible, but also constrained to ensure the success of the narrative goals. The model allows to combine scripted and planned behavior. The state of the story goals is continuously reflected to the narration module to allow it a fine-grained control over the development of the story.

1 Introduction

We report on research conducted in the VirtualHuman project[1] that stages a virtual performance involving autonomous characters and allowing human interaction. The initial prototype of our system, first presented at the CeBIT 2004 exhibition, uses a classroom situation involving a virtual teacher, a virtual pupil, and a human pupil. The virtual teacher gives a lesson about the life of stars; both pupils compete to answer the questions of the teacher. Personality traits of the characters can be configured, e. g. a nice or stern teacher, or the intelligence of the pupil influencing the correctness of answers. Our scenario for the second project phase, which is currently being developed, extends the prototype significantly. It revolves around soccer knowledge and involves three virtual characters (a moderator and two soccer experts) and two human users. It has two stages, a soccer quiz and an interactive game where one human user cooperates with one of the virtual experts to create a soccer team to play against a given opponent.

Both scenarios present the interaction in real-time 3D with animated realistic characters that communicate using speech, gestures, and facial expressions. The user can interact with them via speech and mouse input; in the second scenario, an additional infrared camera will allow detection of the presence of human users and resolving pointing gestures. In a setting as described, one must control the

[1] This research is funded by the German Ministry of Research and Technology (BMBF) under grant 01 IMB 01A. The responsibility lies with the author. See also the project web site, http://www.virtual-human.org

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 53–62, 2005.

tradeoff between (a) the dramaturgic goals that have to be honored, (b) the individuality of the behavior of the characters, and (c) the fact that the actions of a human user may not always be directed towards the story goals, or even counter-productive. It may also be that generated behavior, while achieving the plot goals *per se*, is not dramaturgically pleasing or convincing. There have been some approaches to the problem of generating actions for virtual characters in such a setting, mainly resulting in severely limiting the character's options (see e. g. [7]), or controlling the environment to ensure the user is unable to wreak havoc (e. g. [11]). In our system, we strive to keep autonomy and control balanced by coupling the characters with a narrative engine that takes the authority of a director. Scripted behavior based on state automata is useful for static parts of a performance, like a cut-scene or a fixed introduction. However, to achieve an interactive performance that is diverse and not too predictable, we need to use a more flexible behavior generating mechanism. For this, we integrate scripts with the plan-based dialogue engine originally developed for the SmartKom system.

This paper focuses on the dialogue model for high-level dialogue actions and action planning for the individual characters. Some important aspects of the interaction that are addressed in our system are therefore not discussed here. A Fusion and Discourse Engine (FADE) does fusion of multi-modal input and represents it in a modality-independent representation, and a multi-modal generator devises the actual output actions to be rendered, so action planning does not need to deal with modality issues. On other processing levels, our system supports generating convincing reactive and semi-conscious behavior described in [6]. Issues of turn-taking and synchronizing the verbal and non-verbal interactions in real-time omitted here are described in [10]. We also do not go into dialogue history representation, which is handled by the FADE module, too.

The model and implementation described here are not meant to be sufficient to render a meaningful interactive performance in itself. An active counterpart is required, a *Narration Engine* (as e. g. described in [2]) that models and pursues the story progress, providing high-level goals for the characters: *what* is to be achieved under which constraints. Given this, the action planner should find a way of *how* to do it, realizing the details of the interaction. The next section describes the abstract model used to represent the dialogue structure in our system. Section 3 then talks about the framework implementing the model, and gives a small example. We conclude with an outlook on future planned work.

2 The Dialogue Model

The structure of the dialogue is represented in three hierarchical levels: Dialogue *acts* are used as the atomic units of communication between interlocutors, dialogue *games* specify rule-governed exchanges of dialogue acts, and *activities* use series of dialogue games to implement behavior that is directed to achieve the goals of the characters[2]. One goal is that instances of the elements on different levels can be generalized and re-used as basic building blocks for storytelling.

Dialogue Acts. The set of dialogue acts is shared and agreed upon between all interlocutors. The set of useful acts depends on the particular scenario; we borrow and extend the basic structure from commonly used dialogue tag sets (e. g. [1]). However – at least for the time being – we do not aim to be complete or sound from the speech-theoretical point of view; rather, we use these acts as a starting point to assemble the interaction types we need for our scenario. Dialogue acts specify semantical content to be conveyed (e. g. the focus of a question) as well as preconditions and postconditions. Preconditions must hold for an an act to be usable, while postconditions are assumed to hold afterwards. This is not guaranteed for a character, however, but only a subjective assumption: characters can have several versions of dialogue acts with added private conditions to the dialogue act in their repertoire. While a character A may try to use a *thank* act to improve the social relations to character B, a “grumpy” B might well choose to ignore this attempt depending on its own mental state.

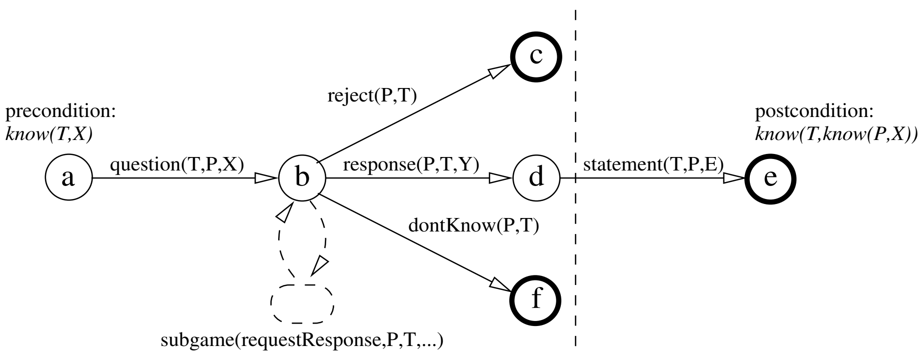


Fig. 1. Example composite game involving a teacher (T) asking a question about X to a pupil (P) and evaluating the answer Y giving an evaluation statement E. Terminal states are in bold. At state b, P could embed a subgame to ask a counter-question.

Dialogue Games. The notion of dialogical interaction seen as a game has a long history (see e. g. [3], chapter 5.2.1, for an overview). In our approach, dialogue games are used to structure and coordinate the exchanges of dialogue acts between the interlocutors, and to generate reasonable expectations for future utterances. As is commonly the case with games, ours consist of sequences of *moves* along with *rules* stating constraints on legal moves for a (dialogue) participant to make in a given situation. To make a move, a participant sends

[2] We focus on dialogue here, and almost all of the interaction in our system is dialogical; however, physical acts such as moving an object and subsequently receiving perceptions about the effect can also be seen as a “dialogue with the world” and we treat them as such in our system.

an instantiation of a dialogue act to another participant. Such a game can be depicted as a finite state automaton. Fig. 1 shows an example which we will use as an illustration. It is a composite game created by the composition of a request-response game and a statement dialogue act (left and right of the dashed line). Composite games are created by sequencing, embedding, iterating or parallelizing other games and dialogue acts – cf. [8] – but not all terminal states of a game need to get a transition to an appended successor. In the example, the statement is only added as a successor to the state d in the original game, which was a terminal state before. A composition can and often will override conditions for the game states of the composite parts, or add new ones. In this case, the original request-response game has a precondition that the initiator does not know the fact asked for, and a postcondition that it assumes to know it afterwards. These conditions are erased for the teacher-pupil question game.

Like the dialogue acts, the basic game set is shared knowledge (the games can also be seen as "social conventions"), but characters may add private preconditions and postconditions and have several versions of a game for different conditions. Even though the shared knowledge about dialogue games is used as a device for the participants to pursue joint goals (cf. [4]), the caveat that postconditions are not guaranteed to hold after a game is even stronger than in the case of a single dialogue act, since there generally is more than one way for a game to terminate. While the example game has the postcondition that T will know whether P knows the answer to X, this is only the case if the game terminates with state e, not in c or f. Thus, there must be a way to choose between possible transitions, and to predict other's choices. When there is a choice between several moves during a game, the character making the turn examines the (shared and private) preconditions of the moves, and selects one that fits best to the current situation. A character predicting the move of another can use the same method using its own version of the game. Conditions can be Logical requirements or context conditions. Logical requirements are checked against the private world state of the character. For example, a *greeting* move can be prohibited if the character remembers that it has already greeted the potential addressee. Context conditions come in two flavors. A precondition can use the static and dynamic *character traits* to constrain the choice of a move. This way, a character can refuse to answer a question from a character he is (currently) hostile towards even if he knows the answer. *Game parameter* conditions check the content of moves, e. g. whether the content type of a question is related to soccer or not. Additionally, a *probability weight* can be used to assign a choice relative weight, introducing an element of chance in the interaction.

A move is selected from a set of conditional alternatives as follows: All choices with unfulfilled logical requirements or character trait conditions get removed. This set is reduced to its members with the most specific applicable parameter conditions. Finally, a random selection is made between the remaining choices, where the chance of a member is proportional to its relative probability weight.

Activities. On the top level of the dialogue model, characters are executing *Activities* to achieve the goals of the narrative. They use preconditions and postconditions like the dialogue games, and also can exist in several versions with different condition sets. Again, it is not guaranteed that an activity will succeed, since it would require complete certainty about the knowledge and cooperativity of other participants. Activities achieve their communicative tasks by using single dialogue acts or playing dialogue games, or delegate the task to sub-activities that can be selected based on preconditions as described in the previous subsection. The model treats activities as black boxes that can be used to perform any kind of computation necessary to determine what games or acts to use. In its simplest form, an activity can just use a fixed sequence of acts or follow the rules of a game, which then results in scripted behavior.

3 The Framework

The dialogue manager of VirtualHuman comprises a system of interacting modules called *CDEs* (Conversational Dialogue Engines). There are three kinds of agents that interact in a CDE system: The human user, the virtual characters, and an interface to the narrative control module. From the point of view of each agent, it is transparent what kind another interlocutor belongs to. We achieve this by using different types of CDEs that, while operating differently internally, communicate in a uniform way. Destinations and sources for communication events are specified abstractly by way of communication *channels*. A CDE can publish or subscribe to any number of input and output channels. The messages are in a uniform format that exchanges *Act*s derived from entities defined in the system ontology, of which dialogue acts are a subtype. Channels that pass messages from and to external modules that use another message format (e. g. the Narration Engine) implicitly translate between the formats.

Character CDEs perform activities, games and atomic actions to achieve narrative goals, as outlined in the previous section. They perceive changes in the

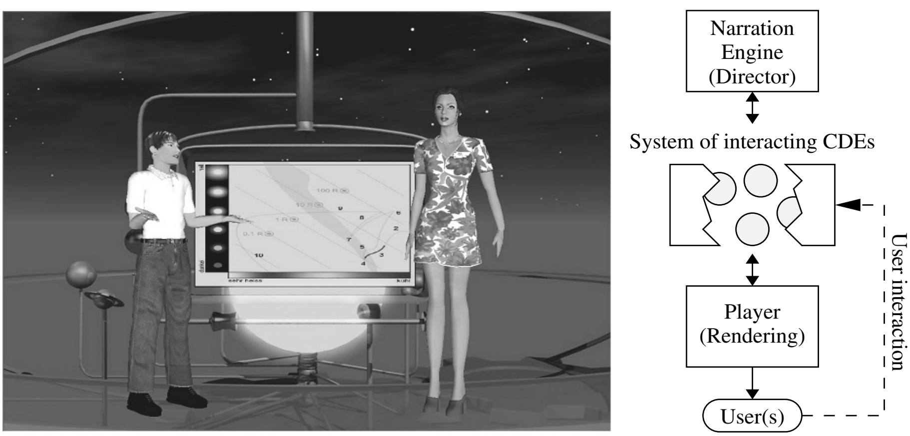


Fig. 2. Screenshot of the teaching scenario and toplevel control flow in the system

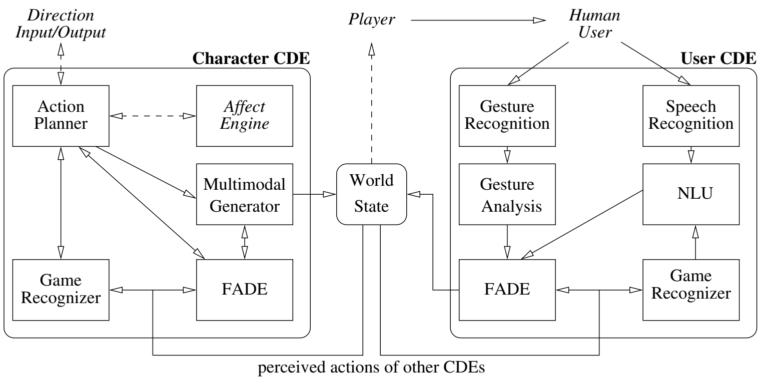


Fig. 3. Internal components of character CDEs (left) and user CDEs (right). Dashed lines indicate translations to external message formats.

environment effected by other characters. The contributions of human users enter the system by *User CDEs* containing multi-modal recognizers and interpreters. One *Director CDE* connects to the Narrative engine and embodies the (invisible) director of the scene. Upon a scene change, it sets up the stage by instructing the player about the objects present and camera positions, and it adds and removes other (character and user) CDEs in the system according to the scene cast.

3.1 Cognitive and World State Modeling

The knowledge foundation of our system is an ontological representation of the objects and relations of the domain as well as the dialogue processing itself. We extended a base ontology introduced in [12] modeled in the Protégé ontology format. They represent all physical and representational objects known to the system, including the building blocks of the dialogue model. The current state of the story world is stored in the world state as typed feature structures representing instances of the ontological objects, along with facts and relations over these objects. The individual character CDEs can obtain – perceive – this information, but the CDE controller may impose perception restrictions. For example, a CDE may be prevented from perceiving an object that is not in the "line of sight" of the character, just as a human user might not see an obstructed object. Perceived objects, facts and relations are added to the subjective world state of the character. The private world state also stores other information private to the character, such as the list of private beliefs, and the character model. A character model describes static character traits and a dynamic representation of the private *affective state* that influence character behavior. The affective state is dynamically updated by a module called *AffectEngine*, which computes a multidimensional value based on the psychological "Big Five" model (a linear combination of personality trait quantifications for *openness*, *conscientiousness*, *extraversion*, *agreeableness*, and *neuroticism*) and derived values for anger, happiness, and so on. Perceived utterances are used by the AffectEngine to update the model, e. g. increasing the anger value when the character is insulted.

3.2 Narrative Control of Character Actions

A two-way communication protocol using the language directionML allows the Narration Engine control over the actions of the characters. It sets narrative goals for them to achieve and receives back information about their success or failure and changes in the state of the world. [2] describes an earlier version of directionML without invariants and return value specifications. We want to divide the work of ensuring the story progress and controlling the details of character actions between the Narration Engine and the Action Planner of the respective CDE. It is not easy to determine the amount and type of information necessary to ensure a smooth progress of the story.

The Narration Engine can start and stop an activity any time, and request feedback about its state. DirectionML additionally supports constraining an activity, or requesting information about a CDE's internal state. A simple constraint is giving a *timeout*, after which the goal fails automatically. Activities can also be constrained by *invariants* (predicates on role values) that must not be violated during execution. In our teaching example, the teacher character can be instructed to abort questioning after three false answers. The Narration Engine can then adapt the story and decide that the teacher should become angry, or should take a detour to explain things again more carefully before continuing the lesson. The most versatile possibility is to request the current value of any role used in an activity (e. g. the exact answer that was given by the pupil), and to adapt the story based on that.

3.3 Implementation

The method used to determine character actions has far-reaching consequences. Two opposite ends of the spectrum of established methods would be the *state-based* and *plan-based* approaches. State-based approaches are often used for dialogue systems with simpler domains and an overseeable set of possible dialogue situations. When the scenario grows more complex, the number of necessary states and transitions can however become very large and hard to maintain. On the other hand, plan-based approaches use a logic description of the domain to devise sequences of operator applications that lead to desired world states. They are powerful and flexible, but it is also hard to control the computational demands of a full-fledged planner operating over a nontrivial domain.

Our approach combines both paradigms. When a CDE is assigned a goal by narrative control, a corresponding activity is started. Activities share a common interface, which specifies that an activity (a) registers with the CDE the signature of the goal it implements (i. e., the ontological object representing the goal and the set of formal parameters (roles) that can be assigned (required or optional), (b) accepts being invoked with parameters matching its signature and the possible narrative constraints, i. e., timeouts, invariants and return value specifications, and (c) guarantees to provide the narrative feedback according to the narrative constraints. There are four types of activity implementations:

- *Game* activities have as signature a dialogue game and role values (minimally the other dialogue participant). They execute the dialogue game, respecting the preconditions as described in Sec. 2, and effect the assertions given in the postconditions of the performed moves.
- *Planner* activities take a parameterized goal and invoke an external planner (JSHOP2) to determine a sequence of other activities to achieve the goal. They succeed only if such a sequence is found and return it as their result. A hardcoded activity can use a planner sub-activity to dynamically find a sequence of game activities to satisfy a set of postconditions, and then execute them as sub-activities.
- *Distributor* activities choose from a set of sub-activities matching their own signature, but with different preconditions, and delegate their goal. For the choice, the method from Sec. 2 is used.
- *Hardcoded* activities are pure Java classes that implement the interface, and can do any computation in Java. They can be used to render fixed cut scenes, or to implement behavior that is not yet capturable with the other activity classes. Implementation patterns reoccurring here are analyzed to be integrated in a general activity specification language (see conclusion).

Activities are realized as concurrent processes executed by a CDE, with direct access to its private world view to read and write assertions, and can start sub-activities that inform their parents about their preconditions when they start to execute. Initially, a character CDE spawns one *root activity* that receives all incoming acts. The root activity can initially start sub-activities executing in the background. These are called *internal* activities, since they do not require a narrative goal. They generate behavior for internal character goals, e. g. a desire to automatically greet perceived other characters. Activities triggered by narrative control are called *external* activities. An activity is also started when a dialogue move from another dialogue participant arrives, and the CDE joins in a dialogue game. When a game starts, the initial move together with the game characterization is sent to the root activity. If a received dialogue game matches the preconditions of a sub-activity, the game is delegated to the sub-activity (and so on recursively), otherwise the activity tries to deal with it itself. Subsequent acts belonging to a delegated game are directly sent to the corresponding activity.

A Short Example. A sample interaction, as realized in our first demonstrator, proceeds as follows: after setting up a scenario with a virtual teacher and pupil and one human pupil, the NarrationEngine starts a *TeachQuestion* activity for the teacher to ask teaching questions about some subject (an ontological object), and a *Listen* activity for the virtual pupil (Sven). *TeachQuestion* selects from a set of available questions on the subject and available addressees. It initiates a question-response-evaluation game (see Fig. 1), asks the question, e. g. to the human, and waits for an answer: *Teacher: "The brightness of stars here depends* *on two parameters, which are... ?"* Sven's *Listen* overhears the question via the world state, but first does not answer: Sven was not addressed. However, Sven is rather bold: after a timeout, Sven's *Listen* activity takes the initiative and contributes an answer: *Sven: "Easy! Size and distance from Earth"*. This answer also is colored to reflect an exuberant mood of Sven. *TeachQuestion* accepts it as a possible continuation of the game with a changed addressee, and proceeds with an evaluation of Sven's answer, completing the game: *Teacher: "Okay, good answer!"* How an interaction goes depends on numerous parameters. Sven's tendency to barge in depends on his static and dynamic character traits and his knowledge base, as described in Sec. 3.1. The correctness of Sven's answer depends on chance influenced by a static "intelligence" trait, and influences the dynamic mood of the teacher. In case nobody answers before a timeout, another way to get the subject across could be chosen by the NarrationEngine (e. g. an interactive experiment). If the user answered quickly enough, the game would complete without Sven's intervention. Small differences can cumulate and lead to different outcomes of the story.

4 Conclusion and Future Work

We presented an action planning approach for multi-modal dialogue systems involving multiple characters and users that can be combined with a narrative module to enact a controlled story with complex planned and scripted interactions. What the narrative module can effect with the system directly depends on the capabilities of the characters; therefore, the action planning mechanism must be taken into account in the development of the story and world representation. As stated, the VirtualHuman system is still under development (a demonstrator of the final system is planned for CeBIT 2006). The astronomy lesson of our first prototype, which takes around 10–15 minutes, had a first evaluation involving 19 pupils (see [5]). The system was perceived as interesting and entertaining, albeit still lacking in broadness. The possibility of multi-modal interaction was appreciated, but criticized for reaction times and little flexibility of speech recognition, which we will address in the second phase of the project. Also, there was a desire for a more versatile understanding of spontaneous questions.

Our current work focuses on the ontological content for the extended second scenario and the development of a dedicated activity language to allow for easy high-level activity specifications to close the gap between using pure and planned dialogue games and activity coordination implemented directly in Java. It looks promising to let such a language be inspired by e. g. the agent behavior language ABL described in [7]. We also need to further investigate which information is useful for the Narrative Engine to determine the course of the story, and which constraints can be imposed on the characters to make them true to the story while remaining sufficiently autonomous to be interesting. To take into account the changed scenario, and to improve the eloquence of the characters, the set of dialogue games will be further extended to include elements common in quiz and soccer discussions, such as rhetorical questions, expert quibbles, etc.

Another interesting possibility in the performance, e. g. for games, would be to allow for live interchangeability of human users and virtual characters during the story, i. e., that a human could take over the role of a virtual character on-the-fly during the interaction, or let a virtual character continue to play her own role. Apart from the affected CDE that would have to make a live transition (including the dialogue history) from user to character CDE or vice versa, the rest of the system would not need to be disturbed. It could also be useful to enable the characters to learn from experience and adjust the transition probabilities of dialogue games accordingly. This way, an attitude like *"Fred does not know anything about soccer!"* could be developed dynamically during the interaction and influence the course of the action, e. g. causing a character to prefer asking somebody else if a question is about soccer.

References

1. Jean Carletta, Amy Isard, Stephen Isard, Jacqueline Kowtko, Gwyneth Doherty-Sneddon and Anne Anderson: HCRC Dialogue Structure Coding Manual, Universities of Edinburgh and Glasgow (1996)
2. Stefan Göbel, Oliver Schneider, Ido Iurgel, Axel Feix, Christian Knöpfle, and Alexander Rettig: VirtualHuman: Storytelling and Computer Graphics for a Virtual Human Platform, Proceedings of TIDSE04, Darmstadt, Germany (2004)
3. Joris Hulstijn: Dialogue Models for Inquiry and Transaction, PhD thesis, University of Twente, Sweden (2000)
4. Joris Hulstijn: Dialogue Games are Recipes for Joint Action, Proceedings of Gotalog Workshop on the Semantics and Pragmatics of Dialogues, Gothenburg, Sweden 2000
5. Marion Langer, Frank Hülsken and Christian Eckes: Evaluation des T24-Demonstrators im BMBF-Forschungsprojekt "Virtual Human", Technical Report 3, Fraunhofer-Institut für Medienkommunikation, St. Augustin (2005) (unpublished)
6. Markus Löckelt and Norbert Pfleger: Multi-Party Interaction With Self-Contained Virtual Characters. Proceedings of the DIALOR workshop on the semantics and pragmatics of dialogue, LORIA, Nancy, France (2005) pp. 139–143
7. Michael Mateas: Interactive Drama, Art, and Artificial Intelligence. PhD Thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA (2002)
8. Peter McBurney and Simon Parsons: Games that Agents Play: A Formal Framework for Dialogues Between Autonomous Agents. Journal of Logic, Language and Information, 11(3):315–334 (2002)
9. Peter McBurney and Simon Parsons: Dialogue Games in Multi-Agent Systems, Journal of Informal Logic, Special Issue on Applications of Argumentation in Computer Science (2002)
10. Norbert Pfleger and Markus Löckelt: Synchronizing Dialogue Contributions of Human Users and Virtual Characters in a Virtual Reality Environment. Proceedings of the 9th European Conference on Speech Communication and Technology (INTERSPEECH), Lissabon, Portugal (2005, to appear)
11. Mark O. Riedl and R. Michael Young: Character-Focused Narrative Planning for Execution in Virtual Worlds. Proceedings of the 2nd International Conference on Virtual Storytelling, Toulouse, France, (2003) pp. 47–56
12. Stuart J. Russell and Peter Norvig: Artificial Intelligence: A Modern Approach. Prentice-Hall, Upper Saddle River (1995)

An Emotional Architecture for Virtual Characters

Ricardo Imbert and Angélica de Antonio

Facultad de Informática, Universidad Politécnica de Madrid,
Campus de Montegancedo, s/n,
28660 Boadilla del Monte, Madrid, Spain
{rimbert, angelica}@fi.upm.es

Abstract. This paper presents the mechanisms proposed by a generic cognitive architecture for virtual characters with emotional influenced behaviors, called COGNITIVA, to maintain behavior control at will without giving up the richness provided by emotions. This architecture, together with a progressive specification process for its application, have been used successfully to model 3D intelligent virtual actors for virtual storytelling.

1 The Role of Emotions in Virtual Storytelling

Traditionally it was considered that intelligent behaviors could only be produced from pure rational reasoning processes. However, everyday experience shows that pure rationality fails when trying to explain many human behaviors, in which the emotional component has a decisive weight. Although quite a lot of efforts have been made to consider emotions in computational models of cognitive processes, emotion is still perceived by many as a non-desirable quality for a computational system [1]. Emotion, mood, personality or attitudes have been considered synonyms of loss of control and entropy growth, and, as such, non-desirable qualities and even something "non scientific" [2].

On the opposite side, recent theories [3] [4] suggest that emotions are an essential part of human intelligence, and are of paramount importance in processes such as perception, learning, attention, memory, rational decision-making and other skills associated to intelligent behaviors. Even more, it has been stated that an excessive deficit of emotion may be harmful to decision-making [5]. Emotion is essential to understand human cognition.

Many of the behaviors produced from pure rational models are far away from those observable in human beings, specially when these models are applied to contexts such as storytelling. Taking emotions away from most stories would make them boring and unbelievable. It is known by everyone that the characters in most stories sometimes do not make the most reasonable decisions, and many times their behavior is strongly influenced by their emotional state. If we want to be able to build believable virtual storytelling, reason and emotion should not be considered as antagonistic and irreconcilable concepts, but as complementary strengths that act upon the mental processes of our virtual actors.

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 63–72, 2005.

2 A Cognitive Architecture to Manage Emotions

Our proposal is to model actors using a truly emotionally-oriented architecture, not a conventional architecture with an emotion component. In our opinion, explainable and elaborated emotion-based behaviors can only emerge when the whole architecture has an emotional vocation.

The architecture that we propose, called COGNITIVA, is an agent-based one. Agents are a common choice to model autonomous virtual actors (not necessarily human-shaped actors but also animals or fantastic creatures), since they present a structure and operation suitable to their needs. Considering an agent as a continuous *perception-cognition-action* cycle, we have restricted the scope of our proposal to the "cognitive" activity, although no constraint on the other two modules (perceptual and actuation) is imposed. This is the reason why this architecture will be sometimes qualified as "cognitive".

In COGNITIVA, emotions are not considered just as a component that provides the system with some "emotional" attributes, but all the components and processes of the architecture have been designed to deal naturally with emotions.

Along this paper, we will analyze how COGNITIVA deals with emotions. Every component and function will be exemplified through the application of the architecture to the modeling of some characters in a well-known fairy tale, Little Red Riding-Hood.

COGNITIVA is a multilayered architecture: it offers three possible layers to the actor designer, each one corresponding to a different kind of behavior, viz reactive, deliberative and social (see Fig. 1). The interaction of these three layers with the other two modules of the actor, the sensors (perceptual module) and the effectors (actuation module), is made through two specific components, the interpreter and the scheduler, respectively.

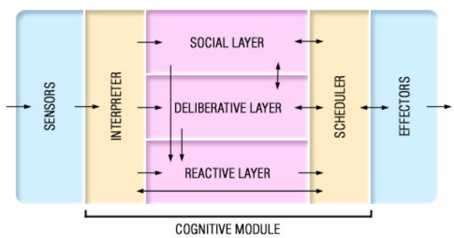


Fig. 1. General structure of COGNITIVA

The cognitive module described by COGNITIVA receives from the perceptual module (the actor's sensors) *perceptions* of the environment. This input may not be directly manipulable by most of the processes of the cognitive module and must be interpreted (for instance, sensors might provide measures about light wavelengths, but the cognitive module could only be able to manage directly colors). In other situations, many inputs may be irrelevant for the actor, and should be filtered (for example, when the woodsman is grabbing the wolf who just ate Little Red Riding-Hood, he would not mind anything about Grandma's house decoration).

COGNITIVA provides a component, called **interpreter**, which acts as an interface between sensors and the rest of the cognitive module, receiving the perceptions coming from the perceptual module, filtering and discarding those non-interesting to the actor, and translating them into *percepts*[1], intelligible by the rest of the components and processes of the cognitive module.

On the other hand, each time some internal reasoning process, in any of the three layers of the architecture, proposes an action to be executed, it must be properly sequenced with other previous and simultaneous action proposals. This should not be a responsibility of the actuation module (effectors), since this module should not need to have any information about the origin and final goal of the actions it receives. COGNITIVA proposes a component, the **scheduler**, to act as interface between the cognitive module and the effectors, managing an internal agenda in which action proposals are conveniently sequenced and ordered. The scheduler organizes the actions according to their priority, established by the reasoning process that generated them, and their concurrence restrictions. Once it has decided which is/are the most adequate action/s to be executed, it sends it/them to the effectors.

The dynamics of the architecture follow a continuous cycle, represented in the Fig. 2, that leaves no room for chaotic behaviors.

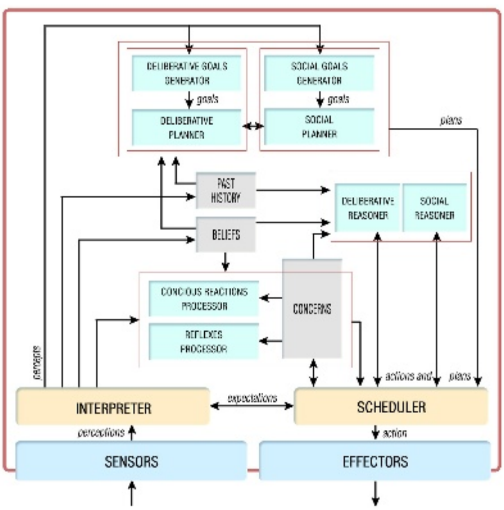


Fig. 2. Internal components and processes of COGNITIVA

[1] Name proposed by Pierce [6], in the context of visual perception, to design the initial interpretative hypothesis of what is being perceived.

3 What Does an Emotional Virtual Character Know?

3.1 Beliefs About the Current State of the World

Beliefs represent the information managed by the actor about the most probable state of the environment, considering all the places, objects and individuals in it. Among all the beliefs managed by the actor, there is a small group specially related to the emotional behavior. This set, that has been called the actor's *personal model*, is composed by the beliefs that the actor has about itself. More precisely, this personal model consists on personality traits, moods, physical states, attitudes and concerns.

COGNITIVA defines a taxonomy of beliefs, depending on their object and their nature. On one hand, a belief may refer to a *place* in the environment, to *objects* located in the environment, and to other *individuals*. Besides, the agent maintains beliefs concerning the *current situation*, for instance, a belief of Little Red Riding-Hood about the current situation may be the fact that she is engaged in a conversation with the wolf. That is not information about the wolf, nor about Little Red Riding-Hood, but about the situation that is taking place.

Beliefs about places, objects and individuals may include:

- **Defining characteristics (DCs)**, traits that mark out the fundamental features of places, objects or individuals. DCs will hardly change in time, and if they do, it will happen very slowly. For instance, the distance from Little Red Riding-Hood's home to the grandmother's house and their location may be DCs about the virtual stage (place); the capacity of her basket is a DC for an object; and a DC about her grandmother (individual) is its name.

  Among all the DCs that an actor can manage, COGNITIVA prescribes the existence of a set of *personality traits* (P) for individuals. Personality traits will mark out the general lines for the actor's behavior. For instance, the wolf can be provided with two personality traits, *ferocity* and *dishonesty*.
- **Transitory states (TSs)**, characteristics whose values represent the current state of the environment's places, objects or individuals. Unlike the DCs, whose values are, practically, static in time, the TSs values have a much more dynamic nature. Some examples of TSs could be the *number of characters on stage* (TS of a place), the *content* of the basket (TS of an object), or the *position* in the scenario of the wolf (TSs of an individual).

  COGNITIVA considers essential two kinds of TSs for individuals: their *moods* (M), which reflect the emotional internal state of the actors; and their *physical states* (F), which represent the external state of the actors (the state of their *bodies* or representations in the virtual environment). In our fairy tale, Little Red Riding-Hood could have as moods *happiness*, *fear* and *surprise*, and the wolf could have as physical states, *hunger* and *fatigue*.
- **Attitudes (As)**, which determine the predisposition of the actor towards the environment's components (places, objects and individuals). Attitudes are less variable in time than TSs, but more than DCs. Examples of attitudes selected for our scenario are the *confidence* of Little Red Riding-Hood towards the wolf, as well as her *appreciation* towards Grandma. Attitudes are

important to guide the actor's decision making, action selection and, above all, to keep coherence and consistency in the actor's interactions.

The elements of the personal model in our architecture have been modeled with a fuzzy logic representation. Fuzzy logic linguistic labels are nearer to the way in which humans qualify these kind of concepts (it is usual to hear "I am *very* happy", instead of "My happiness is *0.8*"). Besides, fuzzy logic is a good approach to manage imprecision.

Relationships among personal model elements are a key point in COGNITIVA. Many of these beliefs are conceptually closely related, and have a direct influence on each other:

– Personality traits exert an important influence determining emotions. For instance, in a similar situation, the value of the mood *fear* will be different for a *courageous* woodsman than for a *pusillanimous* one.
– The set of attitudes of an actor has some influence on the emotions that it experiences. For instance, the sight of the wolf in Grandma's bed will produce an increment on the woodsman's *fear*, because of its attitude of *apprehension* towards wolves.
– Personality traits, in turn, have influence on attitudes. The *apprehension* towards the wolf will be different depending on the value for the personality trait *courage*: a cowardly woodsman will feel absolute rejection towards wolves, whereas a courageous one just will not like them.
– Physical states have also influence on emotions. For instance, when the wolf is *hungry*, its *happiness* will decrease.
– Finally, personality traits exert some influence on concerns. This influence will be explained later on.

All these relationships have been designed and implemented through special fuzzy rules and fuzzy operators. The result is a set of fuzzy relationships, which might include the following:

```
courage DECREASES ⟨much⟩ fear

courage DECREASES ⟨few⟩ apprehension

apprehension DECREASES ⟨some⟩ happiness

apprehension INCREASES ⟨much⟩ fear
```

3.2 Knowledge of Past Events

Behaviors that do not take into account past events are disappointing to human observers, specially in storytelling. COGNITIVA considers two mechanisms to maintain the actor's past history information:

– **Accumulative effect of the past:** this is an implicit mechanism, related to the way in which beliefs are managed. External changes in the environment or internal modifications in the actor's internal state may produce an update of the actor's beliefs. In the case of transitory states, this update is performed as a variation —on higher or lower intensity— on the previous value of the belief, avoiding abrupt alterations in the individual's state.

– **Explicit management of the past state:** an accumulative effect of the past events may not be enough to manage efficiently the past state, because it does not consider information related to the events themselves or to the temporal instant in which they took place. COGNITIVA maintains explicit propositions related to any significant event —to the actor— that happened. In the wolf's past history we could find events such as *talked to Little Red Riding-Hood*, *ran to Grandma's house*, *ate Grandma*... Past history allows the actor to reason considering facts occurred in past moments. As a possible way to implement it, an inverse delta based mechanism has been developed to manage past events.

4 How does an Emotional Virtual Character Behave?

Emotions, in particular moods, may be a strong force to drive the actor's behavior. As it was seen before, emotions are part of the state of the agent. If their values are properly updated and their effects are reasonable, the outcomes of the emotionally based behavior will not be unpredictable, but coherent responses.

4.1 The Effect of Perceptions on Emotions

COGNITIVA provides some mechanisms to update and control the internal state of the actor and, in particular, to control the values of the components of the personal model.

In the first place, the interpreter will direct the interpreted percepts to the convenient processes in every layer of the architecture. The interpreter also feeds information for updating past history and beliefs. Most of that updating may be more or less automatic, and needs no further processing. For instance, when Little Red Riding-Hood perceives her Grandma in bed, the interpreter will update automatically the beliefs about her appearance (size of her ears, eyes and teeth, for instance).

However, that is not the case for moods, and moods are the core of emotional behavior. The new value of moods depends on their old value and on the perceptions, but also on what was expected to happen and to which degree that occurrence was desired. Moods need a new factor to be conveniently generated. With this aim, COGNITIVA includes the mechanism of the **expectations**, inspired on the proposal of Seif El-Nasr [7], which has been adapted, in turn, from the OCC Model [8].

Expectations capture the predisposition of the actor toward the events — confirmed or potential. In COGNITIVA, expectations are valuated on:

– Their *expectancy:* Expressing how probably the occurrence of the event is expected.

– Their *desire:* Indicating the degree of desirability of the event.

Through expectations, the interpreter has a mechanism to update moods from perception:

– When the event **occurrence has not yet been confirmed**. Moods will be updated depending on the degrees of expectancy and desire for the event. For example, if the Grandma knows that Little Red Riding-Hood was going to visit her, but it is getting late, she may elaborate the expectation "something happened to Little Red Riding-Hood in the forest". That is an undesirable event, whose occurrence has not been confirmed yet, that produces a sensation of *distress*, increasing the value of Grandma's "fear" and decreasing the value of her "happiness".
– When the event **occurrence has already been confirmed**. Again, depending on the degrees of expectancy and desire for the event, moods will be updated. For instance, if someone informed Grandma that Little Red Riding-Hood effectively was attacked by the wolf in the forest, her fears would be confirmed, and her *distress* would transform into *sadness*, decreasing considerably the value of her "happiness".
– When the event **non-occurrence has already been confirmed**. The degree of expectancy and desire of the event will determine the updating of moods. For instance, when Grandma hears someone knocking on her door, she believes that Little Red Riding-Hood has arrived safe, so her expectation about Little Red Riding-Hood being attacked vanishes, and *distress* would give way to *relief* by increasing "happiness" and "surprise".

This is how expectations are used to update moods, but, what is the origin of those expectations? Actions will have a set of associated expectations. When the scheduler selects an action to be executed, it informs the interpreter about what is expected to occur in the future, according to that action.

4.2 The Effect of Emotions on Actuation

It is reasonable that actions proposed by the reactive layer have a higher priority in the queue of actions to be executed than those coming from the deliberative o social layers. Even more, it makes sense that deliberative or social action executions are interrupted when a reactive action emerges. Then, does not that mean that reactive, instinctive, passional actions will always take the control of the actuation, leaving out higher level behaviors? Is not that, in fact, losing control? In fact, humans have, to some extent, the ability to control the reactions that logically follow from their emotional state.

The mechanism proposed in COGNITIVA to allow higher level control the actor's actuation is the use of *concerns*. Beliefs represent information about what the actor thinks is the most probable state of the environment, including itself and the rest of the actors. For instance, when the woodsman enters Grandma's house and sees that the wolf has just eaten Little Red Riding-Hood, he will know that he his feeling fear for the wolf, but as this is not at all a desired state for him, he should try to do something to change that state (run away).

Concerns express the desirable/acceptable values for the TSs of an actor anytime, in particular, for **emotions** and **physical states**. Concerns restrict the range of values of the TSs of the actor, expressing the acceptable limits in a certain moment. With this aim, concerns provide two *thresholds*, *lower* and *upper*, for every TS. All the values among them will be considered as desired by the actor; those values out of this range will be unpleasant, and the actor will be inclined to act to avoid them and let them move to the desired range.

Then, a reaction, besides some triggering conditions, the operator to be executed, the consequences of its execution, and some other parameters, such as its priority or its expiry time, will be provided with *justifiers*, i.e., emotional restrictions that must be satisfied to execute the action. Justifiers are expressed in terms of restrictions related to the value of the actor's concerns, that is, restrictions on the desirable state of the actor. For instance, a justifier to trigger a reaction to run away because of the fear produced by the wolf will be:

$$fear > upper\_threshold\_concern(fear) \tag{1}$$

Whenever some actor wants to be able to stand a bit more fear, it first must raise the value of the upper threshold of this concern. If fear does not surpass the value of the upper threshold, the reaction will not be justified and it will not be triggered.

Depending on the personality traits of the individual, which have some influence on concerns as it was mentioned before, the real new value for that upper threshold will be higher or lower. Coming back to the scenario of the fairy tale, if two woodsmen, one courageous and another one easily frightened, enter Grandma's house when hearing Little Red Riding-Hood's cry, and they will perceive the wolf, their fear will raise and a reaction of escape would be triggered. Once they are far enough and their fear has descended under the value of the upper threshold of its corresponding concern, still they will feel worried about Little Red Riding-Hood. However, the new information included in their beliefs, the position of the wolf, prevents them from generating a plan to save Little Red Riding-Hood if they do not consider assuming some risk. As far as Little Red Riding-Hood's life is really in danger, they decide to increase their fear tolerance (the upper threshold of their concern about fear), each one according to their possibilities (their personality traits). They come back to the house, perceive the wolf and, again, their fear raises. But, this time, the level of the fear of the brave woodsman does not surpass its fear tolerance upper threshold, and he grabs the wolf to get Little Red Riding-Hood and Grandma out of its belly. The other woodsman, less courageous, cannot raise enough its fear tolerance upper threshold, and, again, he escapes frightened.

In this way, higher processes of the architecture (deliberative and social) can adjust the value of the thresholds of the actor's concerns to control the instinctive actuation whenever it is not desirable.

5 Related Work

The search for architectures combining rational and emotional behaviors has been a frequent challenge in the last two decades.

Most of the solutions proposed hitherto follow one of two main emotional computational models, generally, the appraisal model (cf. [9], [10]) or the motivational model (cf. [11], [12]). However, although all these models present many interesting virtues, they also suffer from some well-known drawbacks.

Sometimes, emotional elements and mechanisms are so interwoven with the restrictions and particularities of the application context and with the problem faced, that these proposals turn to be very hard to be reused in different contexts. (cf. [13], [14]). In other cases, emotional architectures are very generic, independent from any specific problem (cf. [11], [12]). However, these proposals tend to be less-efficient and computationally demanding. Moreover, they lack of mechanisms suitable to facilitate the adaptation of the architecture to the particular necessities of the problem. Frequently, this adaptation is only achieved through simplification of some of their inherent features and with a lot of effort.

Our solution provides a balance among these two extremes. One of the main characteristics of COGNITIVA—and one of its strengths— is that, having been conceived as a generic, domain-independent architecture, not restricted to any emotional theory or model, and applicable to a wide range of domains and problems, it is accompanied by a progressive specification process in two phases (Functional Specification and Contextual Specification) to be applied for the design and adaptation of the abstract structures and functions proposed, to the particular needs of the application context (the details of the process are out of the scope of this paper, but can be found in [15]).

6 Conclusions

Human and animal behaviors are rarely exclusively explainable through pure reasoning. There exist other emotional factors that influence decisively on them, that must also be considered. However, the efforts until today to build architectures including those emotional factors have not yet succeeded, and emotion is still frequently observed as an undesirable quality to be included in computational systems.

This paper presents some of the mechanisms proposed in COGNITIVA, an architecture with generic mechanisms and structures to build actors with emotionally influenced behaviors, and which is able to deal with the problem of the loss of control in the emotionally based behavior generation.

COGNITIVA does not intend to constitute yet another theory of emotion, but an abstract architecture able to leave room for multiple theories. With this philosophy, in the functional specification proposed we have developed a possible theory (with fuzzy relationships among personality traits and moods, for instance) with the aim of checking the applicability of the architecture, and not the empirical validity of the theory itself —which must not be considered, in fact, as anything else than an example of application.

COGNITIVA has already been applied to quite different contexts and scenarios [16], and it has proved (together with the specification process) to be a useful approach to the construction of virtual characters for storytelling.

References

1. D. N. Davis and S. J. Lewis. Computational models of emotion for autonomy and reasoning. *Informatica (Special Edition on Perception and Emotion Based Reasoning)*, 27(2):159–165, 2003.
2. R. W. Picard. Affective computing. Technical Report 321, MIT Media Laboratory, Perceptual Computing Section, November 1995.
3. J. LeDoux. *The Emotional Brain*. Simon and Schuster, New York, 1996.
4. R. Adolphs, D. Tranel, A. Bechara, H. Damasio, and A. R. Damasio. *Neurobiology of Decision-Making*, chapter Neuropsychological Approaches to Reasoning and Decision-Making, pages 157–179. Springer-Verlag, Berlin, Germany, 1996.
5. A. R. Damasio. *Descartes' Error. Emotion, Reason, and the Human Brain*. Gosset/Putnam Press, New York, 1994.
6. C. S. Pierce. *Collected Papers*. The Belknap Press of Harvard University Press, Cambridge, 1965.
7. M. Seif El-Nasr, J. Yen, and T. R. Ioerger. FLAME — a fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-Agent Systems*, 3(3):219–257, 2000.
8. A. Ortony, G. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge, UK, 1988.
9. C. Elliott. I picked up catapia and other stories: A multimodal approach to expressivity for "emotionally intelligent" agents. In *Proceedings of the First International Conference on Autonomous Agents (Agents'97)*, pages 451–457, New York, 5–8, 1997. ACM Press.
10. A. Staller and P. Petta. Introducing emotions in the computational study of norms. In *Proceedings of the AISB'00 Sympoisum on Starting from Society -The Application of Social Analogies to Computational Systems*, pages 101–112, UK, 2000.
11. S. R. Allen. *Concern Processing in Autonomous Agents*. PhD thesis, Faculty of Science of The University of Birmingham, School of Computer Science. Cognitive Science Research Centre. The University of Birmingham, UK, 2001.
12. D. Cañamero. Modeling motivations and emotions as a basis for intelligent behavior. In *Proceedings of the First International Symposium on Autonomous Agents (Agents'97)*, pages 148–155, New York, 1997. ACM Press.
13. S. C. Gadanho. Learning behavior-selection by emotions and cognition in a multigoal robot task. *Journal of Machine Learning Research*, 4:385–412, 2003.
14. E. de Sevin and D. Thalmann. An affective model of action selection for virtual humans. In *Proceedings of Agents that Want and Like: Motivational and Emotional Roots of Cognition and Action Symposium at the Artificial Intelligence and Social Behaviors 2005 Conference (AISB'05)*, Hatfield, UK, 2005.
15. R. Imbert and A. de Antonio. Using progressive adaptability against the complexity of modeling emotionally influenced virtual agents. In *Proceedings of the 18th International Conference on Computer Animation and Social Agents (CASA 2005)*, Hong Kong, China, 2005.
16. R. Imbert and A. de Antonio. When emotion does not mean loss of control. In T. Panayiotopoulos et al., editors, *Intelligent Virtual Agents*, volume 3661 of *Lecture Notes on Computer Science*, pages 152–165, Berlin, 2005. Springer-Verlag.

Generating Verbal and Nonverbal Utterances for Virtual Characters

Benjamin Kempe, Norbert Pfleger, and Markus Löckelt

DFKI GmbH, Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany
{kempe, pfleger, loeckelt}@dfki.de

Abstract. We introduce an approach to multimodal generation of verbal and nonverbal contributions for virtual characters in a multiparty dialogue scenario. This approach addresses issues of turn-taking, is able to synchronize the different modalities in real-time, and supports fixed utterances as well as utterances that are assembled by a full-fledged tree-based text generation algorithm. The system is implemented in a first version as part of the second VirtualHuman demonstrator.

1 Introduction

Real-time multimodal, multiparty interaction involving humans and virtual characters (VCs) is a challenging effort that has been approached from various directions but still poses problems. Many aspects of multiparty interaction need to be resolved in order to achieve a convincing, believable behavior of VCs. We integrate components originally developed for task-oriented multimodal dialogue systems and virtual storytelling systems. The VCs are modeled as autonomous agents under the control of a narrative "director" component. In this paper, we describe the multimodal generation component used to produce the verbal and nonverbal contributions of the VCs.

The work presented here is conducted in the VirtualHuman (VH) project (see `http://www.virtual-human.org`), which is a long-term research effort aiming at the development of VCs that act as believable, life-like dialogue partners. When the VCs react to contributions of other participants, they need to act with respect to the situational context; e. g. obeying turn-taking conventions. The VCs are guided by *goals*, which they receive from a *narration component*. The user contributions are processed by speech and gesture recognition components and converted to the system-internal abstract ontological representation used for communication among VCs. A 3D player component renders the complete scene for the user.

The scenario of the VH demonstrator is a soccer quiz show with two humans and two VCs as contestants; the host of the show is also a VC. The four contestants are shown short videos of crucial situations in past soccer games. The videos are stopped at the moment of deepest suspense and the contestants have to select one of three possible outcomes. Then the video continues and the contestants who chose the correct answer receive points.

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 73–76, 2005.

To illustrate the capabilities of the generation component of our system, we will use a fragment of the quiz show. We enter the game at the point where the moderator has just shown the first part of the video (Michael Ballack, a famous German soccer player, is just preparing to kick) and is explaining the possible answers:

(1) Moderator: *Well, what will happen next: a)* [counting gesture] *Ballack buries the ball in the corner of the goal. b)* [counting gesture] *the goalkeeper saves the ball with his fingertips or c)* [counting gesture] *Ballack sends the ball to the sky.*

(2) User1: *I think it's "b", the goalkeeper saves the ball.*

(3) VC1: *Hhm,* [gazes at moderator; the moderator gazes back] *I disagree. I think he* [pointing gesture at the video screen where Ballack is still visible] *scores.*

The contributions of the VCs are generated entirely online from abstract representations of their semantic content. (For prescribed, static utterances, like introductions, we also have the ability to use canned text.)

2 Conversational Dialogue Engines

The reasoning for each character is encapsulated in an autonomous conversational dialogue engine (CDE). A CDE contains a working memory (called FADE; cf. [5]), a deliberative unit responsible for action planning, an affective unit modeling the emotional state, and the multimodal generator. Intra-CDE communication is in terms of ontological dialogue acts; externally, messages are exchanged with a narrative unit guiding the interaction towards story goals and the player. Dialogue acts produced by the action planner contain an abstract semantic description of an utterance. The task of the multimodal generator is to transform this into a concrete utterance for the player. Facial expressions showing emotional state are triggered in parallel by the affective unit, adding to the multimodal behavior for the character. The action planning process is described in more detail in [4].

```
<Question>
  <has_initiator> ...
  </has_initiator>
  <has_addressee> ...
  </has_addressee>
  <has_content>
    <ListElement>
      <has_listPosition> ...
      </has_listPosition>
      <has_content>
        <Response>
          <has_content>
            <Parade>
              <has_agent>
                <GoalKeeper> ...
                </GoalKeeper>
              </has_agent>
              <has_style> ...
              </has_style>
            </Goal>
          </has_content>
        </Response>
      </has_content>
    </ListElement>
  </has_content>
</Question>
```

Fig. 1. An example dialogue act

All modules in VH are based on the same hierarchical ontology, which we create by means of the Protégé editor. We represent all information available to VCs as ontological instances. In this hierarchy, the first distinction is between abstract concepts (e.g. representational objects like *Acts*) and events (*Processes*, *PhysicalObjects*, etc.). *NonverbalActs* include gestures like *Iconics* and *Metaphorics* as well as *Gazes*. Properties of these classes may contain information about the meaning, duration and appropriate context. The *DialogueAct* class represents the intention of a character to convey propositional and interactional content. *DialogueActs* are further divided into, e g. *Questions*, *Responses* and *Statements*. Figure 1 shows the dialogue act for answer b) in utterance 1 of the example dialogue. This example contains information about the initiator (the moderator) and the addressees (the contestants) as well as a fragment of the propositional content, representing the enumerated possible answer. The answer itself is represented through an instance of a physical act *Parade* comprising information about the agent (keeper) and style (used body parts or other properties of the action). Our semantically focused representation allows reasoning using underlying concepts rather than just explicit keywords. Since the goalkeeper in example 1 is a person, we address him by name.

3 Multimodal Generation

The generator takes an ontological instance of a dialogue act (see Fig. 1) and turns it into PML syntax (a markup language for communicating with 3D players). The output of the generator contains the spoken utterance, as well as tightly synchronized nonverbal actions (gazes, adaptors, emblematic, iconic, deictic and beat gestures). In the dialogue act, some of the content may be marked as optional and can then be realized, depending on the speaker's emotional state and the discourse context. An utterance will turn out differently, depending on whether an uttered element is a newly introduced concept (*"a car"*), has already been introduced (*"the car"*), etc.

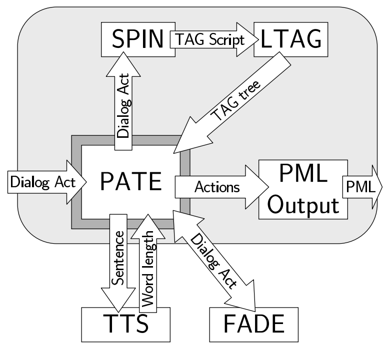


Fig. 2. Architecture of the multimodal generator

Since the generation takes place in real-time, we use several means to cope with time-critical aspects. First of all, generation is bound to be fast and efficient. Secondly, we estimate the amount of time necessary to generate the utterance. If the estimate exceeds a certain threshold, we use additional predefined expressions (like *"hhm"*, *"well"*), suitable in the given situation and mood, to express pensive behavior. Our multiparty scenario asks for a turn-taking approach incorporating general gazing behavior as well as actions to take and yield turns. When the generator detects that its VC is not the one holding the floor, it might attempt (depending on characteristics like urgency and mood) to claim the next turn by making interrupting statements and gestures. Most gestures have to be synchronized with a constituent in the speech utterance. A gesture states the kind of alignment (e. g. starting with a word) as well as the time frame during which it should be performed. Especially in the case of iconics, metaphorics, and emblems, one can identify a gesture's meaning. This gesture should coincide with a constituent of related meaning (e. g. counting gestures are likely to co-occur with an utterance that refers to an enumeration). We encode this meaning and possible accompanying context within the gesture description. Merely attaching gestures to utterances could lead to overlaps. A suitable time spot is sought with respect to the time frame description. Initially, gestures are inserted with minimal duration to enable further attachments. When all gestures are in place, their durations are expanded to maximum appropriate lengths.

The generator is a compound module of four different parts (see Fig. 2). The main control is performed by a production rule system called PATE (cf. [3]). Upon receiving a dialogue act from the action planner, production rules manage initial turn-taking and pensive behavior. PATE also restructures dialogue acts and annotates them with grounding information from conversational context provided by the working memory. The content of a dialogue act is passed to another production rule system called SPIN (cf. [2]). SPIN uses largely generic rules to convert the complex dialogue act into operations on elementary syntax trees for the LTAG generator (cf. [1]). The resulting derivation tree is split into different clauses and words. The sentence is sent to the text-to-speech engine and the word boundaries of its verbalized output are analyzed. This enables the gesture rules, written in PATE, to exactly time the gestures at word boundaries. The library of available gestures is easily extended. In the last step, PATE passes its gesture instances to the PML converter, which describes the whole multimodal utterance in PML code (see Fig. 3) and sends this code to the player. The player renders the presentation for the user.

```
<actions id="ac0">
  <character refId="Host">
    <speak id="s0" dur="2655" ...>
      <text>b) the goalkeeper saves the
        ball with his fingertips </text>
    </speak>
    <animate id="ag0" dur="1400"
        alignTo="s0" alignType="starts">
      <gesture refId="gazeAtVC1"/>
    </animate>
    <animate id="aa0" dur="2000" ...>
      <gesture refId="countTwo"/>
    </animate>
    ...
  </character>
</actions>
```

Fig. 3. Example PML output

References

1. T. Becker: Natural Language Generation with Fully Specified Templates. to appear in: W. Wahlster (ed.), SmartKom: Foundations of Multimodal Dialogue Systems, Springer Verlag, Heidelberg (2006)
2. R. Engel: SPIN: Language Understanding for Spoken Dialogue Systems Using a Production System Approach, Proc. of 7th ICSLP, Denver (2002)
3. B. Kempe: PATE - a Production Rule System based on Activation and Typed Feature Structure Elements, Bachelor Thesis, Saarland University, Saarbrücken (2004)
4. M. Löckelt: Action Planning for Virtual Human Performances, Proc. of Virtual Storytelling 05, Strasbourg (2005)
5. N. Pfleger: FADE - An Integrated Approach to Multimodal Fusion and Discourse Processing. To appear in Proceedings of the Doctoral Spotlight Session of ICMI'05 (2005)

Scenejo – An Interactive Storytelling Platform

Sebastian Weiss[1], Wolfgang Müller[2], Ulrike Spierling[3], and Florian Steimle[4]

[1] Team For Hire, Germany
sweiss@teamforhire.de
[2] Anhalt University of Applied Sciences, Dept. of Computer Science, Germany
wolfgang.mueller@inf.hs-anhalt.de
http://www.inf.hs-anhalt.de
[3] FH Erfurt, Univ. of Applied Sciences, Erfurt, Germany
Spierling@fh-erfurt.de
[4] Univ. of Frankfurt, Dept. of Computer Science, Germany
fsteimle@gdv.cs.uni-frankfurt.de

Abstract. *Scenejo* is an Interactive Storytelling platform, supporting both structured story lines and emergent behavior. Authoring is performed either at the level of a story graph or dialogue patterns. The *Scenejo* platform supports several artificial actors conversing with a number of real actors, representing the users in the system. Artificial actors are visualized as animated 3d characters, and actor responses are presented by speech synthesis in combination with non-verbal behavior.

1 Introduction

The concept of *Interactive Digital Storytelling* has the potential to become a paradigm for future interactive knowledge media. It couples dramatic narrative with interactions of users, providing highest forms of engagement and immersion. It also stands for the connection of games and stories by utilizing inherent structural elements of both. Artificial characters taking the role of actors within a plot play an important role in the concept of Interactive Storytelling.

Digital storytelling agents can achieve more than simply being single virtual guides and virtual tutors, which are commonplace today in a variety of software products. As in stories, their role could be to interact with each other as a set of characters to present a dramatic storyline; and as in games, they have the potential to serve as all sorts of sparring partners for players to interact with, such as representing the bad guys, or companions who ask for help.

Scenejo presents our approach to create a platform for interactive applications in this field. *Scenejo* enables playful simulations of dialogues between several conversational agents and multiple users. It employs animated virtual characters and current chatbot technology as the basis for text-based interactions. The result is an emerging dialogue, influenced by the users' inputs and the bots' databases of possible dialogue lines matching a text pattern coming from either a user or another bot. Bots also take into account parametric settings and scene descriptions provided by authors.

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 77 – 80, 2005.

2 Related Work

Successful implementations of intelligent conversations with animated virtual characters are rare, and there is no real success on the entertainment market to date. One of the few examples examining a middle course between the two approaches of linear stories and emergent behavior is M. Mateas' and A. Stern's Façade [1]. It is based on a specialized dialogue management system and allows users to participate in a predefined and pre-recorded conversation between virtual characters. However, the system's design is focused on a specific scenario and authoring is currently supported for programmers only. The EU-funded project art-E-fact [2] presents a similar integration of simulation and plot. In contrast to Façade, an authoring system is central to the way a story is built in art-E-fact. Storywriters defining digital conversations start with a story graph of explicit dialogue acts, similar to branching, and provide more complex interactions by adding rules and chatbot patterns within nodes of the graph.

With *Scenejo*, we follow a similar goal as in art-E-fact, but start at the opposite end. From the start, we use chatbot text patterns to provide free conversational interaction with users, and in a bottom-up way, we introduce a story graph allowing writers to line up conversational scenes and their parameters.

3 The Scenejo Use Concept

Scenejo is a storytelling platform that renders conversations between multiple artificial characters in combination with free textual user input in a multi user environment. It provides the possibility to let virtual characters enter into a discussion or even a dispute, involving users by asking them for advice, for opinions or by prompting them for directing their interest. The ideal content areas to cover with such dialogues are domains containing either uncertain facts or knowledge elements that are subject to interpretations, such as in the arts or in philosophy – as in its ancient archetype of Plato's Symposium. Another e-learning application field is the training of interactions based on language, or on dialogical skills. This can be suitable in training scenarios for difficult situations or negotiations, as well as simply for mastering the adequate terminology within a knowledge domain.

Scenejo is set up as an authoring system to create and experience emerging conversations. Users create a scenario by defining actors and their properties, such as gender, the voice to be used for speech synthesis, visual representation and a particular dialogue base for each actor.

Scenejo provides a story graph editor to describe the possible converse of a story. Scenes represent the building blocks for the story graph, defining the context and environment for a part of the interactive story. The most important elements of a scene are the actors that participate in this scene. Actors differ in representation, knowledge and vocabulary, and different actors may be assigned to single scenes. Scenes also describe the conditions for an automatic transition to another scene.

When the conversation is started, the bots assigned with the first scene begin to converse, matching text patterns reciprocally. Depending on the pre-authored patterns in the dialogue database, this can result in a rather phased dialogue, a chaotic argument or simply small talk. User input is handled the same way as the bots' patterns.

The dialogue is currently represented being rendered through the animated talking heads (see Figure 1), speaking ad-hoc with a digitally synthesized voice, without any need of pre-recording. Additionally, a text representation of the stage provides an overview about the dialogue from the beginning of a scene up to the current point.

The experience of interacting with the platform of multiple chatbots shows that there is high entertainment value through the fact that the course of the conversation cannot be completely anticipated, even not by the writer of the dialogue patterns. While there are still problems with non-sequitur verbal reactions to user input, people mostly cope with it as within chats in their real life, and as a result, rather assume strange character traits to the bots according to their appearance.

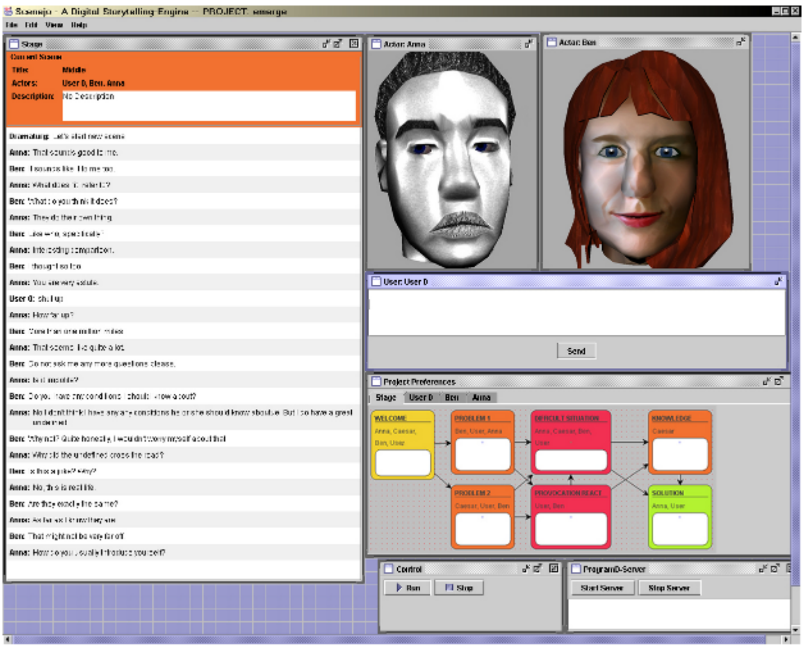


Fig. 1. The *Scenejo* user interface

4 The *Scenejo* Platform

The main components of *Scenejo* are the individual actors and a stage. The stage serves as the central unit coordinating the communication between the actors, managing and interpreting the story graph, and determining the current state of the story. It takes the main responsibility for the course of the story and decides when a transition from one scene in the story to another scene should occur. Scene transitions may be triggered for instance by reaching a specific goal, by the activation of a specific communicative act of an actor, or by the expiration of a timer.

Actors are the acting units in a story and represent the participants in the dialogues. There are two classes of actors: users and artificial characters. It is irrelevant for the stage component which kind of actor it is dealing with. User actors represent the interface between a user and the system, recording the interactions and communicative acts of the user. User interaction is currently limited to chat-like text input via keyboard. In the future this will be extended to speech and video input.

Artificial actors have their own behavior; their reactions to interactions are determined on an individual basis. A decision component is responsible for finding adequate responses to the utterance of another actor. In *Scenejo*, the AliceBot chatbot system [3] is being utilized in this context, where textual input patterns and appropriate responses are described in terms of extended AIML rule sets. System responses have been further extended to RRL expressions to carry additional information on emotions, facial expressions, gestures, and body movements of actors to be shown in the context of an utterance. This information is used in *Scenejo* to provide adequate animated presentations of embodied virtual characters.

Important decisions about responses to utterances are made at actor level. The responsibility for presenting a detailed story graph down to the level of sentences and concrete words is distributed between the stage and the actors. Therefore, the *Scenejo* story graph is usually shallow and broad as in games structures, thus simplifying authoring. The storyline is complemented by another component, the dramatic adviser, which is responsible for managing turn-taking behavior and for granting priority to those communicative acts that are most important for the course of a coherent story. More details on the implementation of *Scenejo* may be found in [4].

5 Conclusions

Scenejo presents an interactive storytelling platform, rendering conversations between multiple artificial characters in combination with free textual user input in a multiuser environment. Actors are visualized as animated 3d avatars with speech output and augmented non-verbal behavior. *Scenejo* presents an approach to the integration of character-based and plot-based approaches.

References

1. Mateas, M., Stern, A. (2003): Integrating Plot, Character and Natural Language Processing in the Interactive Drama Façade. In: Proceedings of TIDSE 2003, Darmstadt, pp. 139-151.
2. Spierling, U., Iurgel, I. (2003): „Just Talking About Art“– Creating Virtual Storytelling Experiences in Mixed Reality. In: Virtual Storytelling, Using Virtual Reality Technologies for Storytelling, Proceedings of ICVS 2003, Toulouse, France, pp. 179-188.
3. ALICE (2005): Homepage of the A.L.I.C.E. Artificial Intelligence Foundation. Online: http://www.alicebot.org (Last accessed: 10.06.2005).
4. Müller, W.; Spierling, U.; Weiss, S. (2005): Synchronizing Natural Language Conversation Between Users And Multiple Agents In Interactive Storytelling Applications, Proc. TESI 2005, Maastricht, The Netherlands.

Drama and Emotion

Did It Make You Cry? Creating Dramatic Agency in Immersive Environments

Janet Murray

School of Literature, Communication and Culture,
Georgia Institute of Technology, USA
janet.murray@lcc.gatech.edu
http://wwww.lcc.gatech.edu/~murray

Abstract. Making you cry is the folk wisdom test for strong narrative involvement, and has long been proposed as a reachable but distant goal for digital games. Immersion is a related phenomenon which has been difficult to achieve in Virtual Reality and Mixed Reality research projects, but reliably achieved in theme park rides by integrating interactivity with strong narrative elements. For Virtual and Mixed Reality environments to achieve narrative power and strong immersion we can adopt the strategies of theme park rides, with special attention to the creation of threshold objects that take the interactor into the virtual space. Spatial presentation of the world, interactions with characters, and differentiating real from virtual events are design challenges without clear genre conventions. Design choices that locate the interactor in a coherent point of view, limit natural language expectations, and conventionalize the boundaries of the virtual world will reinforce immersion and dramatic agency.

1 Story Power and Interactivity

“Can a computer make you cry?”

This was the tag line to the ads announcing the founding of Electronic Arts in 1982, which positioned the company at the intersection of videogames and older storytelling forms. The question assumes that there is something more humanly significant about an entertainment form that evokes tears than one that makes you laugh. It also predicted the change in videogames from a focus on Pacman-like abstractions to one on role playing avatars, adventure heroes, and eventually bourgeois Sim families. Although players were crying over the death of robot companions in text adventure games even before the advent of cinematic videogames [8], EA has yet to prove the point to people like Steven Spielberg who remarked at last year’s opening of the EA Game Lab at the University of Southern California "I think the real indicator [that games are the equivalent of film] will be when somebody confesses that they cried at level 17" [1].

Weeping is an interesting touchstone because it assumes that melodrama is the measure of narrative art. We do not cry at tragedy, as Aristotle pointed out: we react to it with pity and terror and an exalted purgation of emotion. It is our most serious art form because, like poetry, it touches on the truths of life that, in Wordsworth’s phrase,

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 83 – 94, 2005.

"lie too deep for tears." We cry at melodrama, the sentimental form that reaffirms our comfortable expectations of the world while reminding us of poignant loses. Melodrama draws on exaggeratedly good or evil stock characters and familiar plot formulas, and it resolves in happy endings or bittersweet sad ones. Crying at melodrama is a way of feeling sorry for ourselves. In the Victorian era, tears were most often shed over the premature death of adorable children, such as Dickens' Tiny Tim, Little Nell, and Paul Dombey. In current American pop culture we are more likely to cry over television, particularly movies of the week that focus on brave fights against fatal diseases; or news coverage or reality TV shows that focus on victims of disasters or human violence. We do not cry over movies the way we did in earlier decades, but occasionally a romantic tear-jerker like *Titanic* hits upon the right formula to evoke the sentimental tear. Spielberg's work is highly melodramatic, but he usually evokes sympathy and anxious concern for characters rather than tears. We hold our breaths and rout for them to succeed. ET's predicament strikes the note of pathos, but it does not dwell on it: the emotion is channeled into the exhilaration of his mad bicycle ride and successful escape.

Games are very melodramatic but they are more like American movies than they are like Victorian novels. They involve us in intense identification with the outcome of an adventure or the fate of a character, but not usually in the way that makes us weep. The revelation of the monkey holocaust in *Myst* carries a moral and emotional weight that has a real sadness to it. But in most games such massive death would be played out as horror spectacle: a form of melodrama that heightens excitement at the expense of character involvement. Horror movies do not make us cry: they make us scream. Game designers, with their eye on the adolescent male audience, often include tropes from the horror genre. Theme parks also rely upon stock horror visual, auditory, and motion effects to create a predictable and intense user experience. If intensity is the measure of immersion, then researchers might do well to stick to horror conceits.

The *Sims* website is full of user-created stories that make a viewer cry, such as the reenactment of a family member's abusive relationship or a happily ever after courtship story. But the audience creates this intensity out of relatively neutral building blocks. The intrinsic plot of the *Sims* is one of bourgeois success: work hard, make friends in your suburb, keep your house clean, get more stuff. But players imbue the world with other levels of meanings. The death of a family by fire in the Sims can be experienced sad or funny or trivially important depending on how invested the player is in the game. In addition to its sob stories, the *Sims* website also contains stories of gleefully intentional mayhem, little snuff-films with clever ways of doing away with characters.

Tears are an appropriate measure of involvement because they are physiological and suggest authenticity and depth of feeling. Cheers, laughter, and screams are similarly emotive. Screams might be greater evidence of "presence" than tears, but they can be evoked by isolated representations: a looming vampire or a swooping bat would do fine. What do we mean when we ask whether a medium can tell us a story that makes us cry? I think "make us cry" stands for a set of phenomena that do not have to involve actual tears: emotions aroused through identification with the characters; a feeling that the actions of the characters and the events in the imaginary world have dramatic importance, that there is a weight and significance to how things turn out; a suspension of disbelief, in Coleridge's useful term, that lets us enter into and sustain an involvement in a richly detailed imaginary world. We want the characters, events, and settings to fit together in a way that intrigues us, that raises our curiosity about what will happen next, and then rewards that curiosity with a sequence of appropriate yet surprising occurrences We want to care about the fate of the individual characters and see the events in the imaginary world as fitting our deeper sense of how life is. With the most accomplished works of art we want our understanding and emotional knowledge of how life is to be expanded. When we say "a computer will make you cry" we are implying not sad content, but compellingly powerful and meaningful representation of human experience.

Videogames are full of melodrama that arouses a sense of excitement. But they do not necessarily make us cry. We may be able to suspend our disbelief but we do not particularly care about the people or events except as targets or opponents, wins or losses, even though there may be strong story elements at work. In Spielberg's analysis this is because the storytelling is not integrated with the game format:

> "It's important to emphasize story and emotion and character. This is one of the things that games don't do," Spielberg said. "Currently, what games do is they give you the entire story in the run-up to the actual game play from level to level. You get to see a movie, and you're supposed to remember what the stakes are for the characters. But there's no reminder, nothing refreshes who these characters are.
>
> "The next big emotional breakthrough in gaming is being able to tell a story that is consistent throughout the narrative. If the game is 15 levels, it's just like 15 chapters in a story" [1].

Furthermore, he believes that interactivity is at odds with story pleasure.

> "Is the player in charge of the story, or is the programmer in control of the story?" Spielberg asked. "How do you make those two things reconcile with each other? Audiences often don't want to be in control of a story. They want to be lost in your story. They come to hear you be the storyteller, but in gaming it's going to have to be a little bit of both, a little bit of give and take" [1].

Game developers agree that game elements and story elements are often at odds with one another and it is not easy to see how they can be integrated. We can take Spielberg's comment as a potential design principle: **in order to maximize story power we have to integrate the interactor's actions with the story content**. When we are successful we create the experience of **dramatic agency**, the cueing of the interactor's intentions, expectations, and actions so that they mesh with the story events generated by the system [8]. This seamless interactivity is harder to achieve in virtual reality environments in which the interactor may be learning new equipment and operating outside of known conventions of behavior.

2 Story Power and Immersion

In a story-game the narrative power lies in part in the player's immersion in the interactive world. Immersion is an often evoked positive quality of interactive environments, and also of fictional worlds. As I have argued elsewhere, immersion in both circumstances requires clear boundaries between the real world and imaginary domain, encyclopedic detail within the imaginary environment, and transitional objects that help the reader, viewer, or interactor to move from one to another. VR equipment can act as boundary-maker and transitional object. Putting on the equipment or entering the special chamber can act as a passage onto a stage or into a theater audience, triggering a receptive state of mind. We can think of them as **threshold objects** since they take us across a symbolic and sometimes literal passageway [8]. We can take this as another principles of design: **provide the interactor with clear boundaries and with virtual or real objects that take them across the boundary between the real and imaginary worlds.**

The literary critic Marie-Laure Ryan also believes that "interactivity conflicts either with immersion or with aesthetic design, and usually with both." [10] In her view, virtual reality represents a theoretical intersection between the two, but one that is hard to reconcile and that interferes with coherent storytelling. [11] For Ryan, interactive environments are intrinsically non- immersive because the interface interferes with the illusion of the world. She does not believe in a transparent hypertext interface, for example. But she thinks that virtual reality holds the promise of immersion because it is enacted in the body, so that we have immediate reinforcement of the imaginary world without the mediation of reading. We can elaborate this observation into another design principles: **to root the events of the virtual world in physicality of the interactor through expressive gestures, spatialised sound and images, and haptic feedback.**

Brown and Cairns describe immersion as a 3 stage spectrum, ranging from engagement, engrossment, and total immersion, and the barriers to each. They conclude that transparency of interface is particularly important to even minimal immersion and that total immersion is equivalent to Presence, and a fleeting experience. They consider emotional investment and attention to be key elements in intensifying immersive experience, with experiences that demand and reward multiple elements of attention -- visual, auditory, and puzzle-solving -- to be more immersive. [2] We can interpret their taxonomy as a series of design principles. Among the most important is the need **to elaborate the world in a way that demands and rewards attention.**

Games are slowly moving outside the console and PC box to include experiences that are grounded in the body. Dance, Dance, Revolution and PaRappa the Rapper established a new genre that requires the simple interface of a sensor-equipped mat and works on one of the most basic forms of interaction: follow my commands. The Sony Eye-Toy interface games track the user's hand motions without requiring them to wear any cumbersome motion tracking equipment; the interaction is a chase and point action in which the user tries to place their hands to intersect with the fast moving objects on the screen. It may soon be expanded to track head motion as well, which could be used for metagame commands like "rewind" or in game action like peeking around the corner in an attack sequence. Game developers are also looking at mixed reality presentations that capture image of the player and the player's environment and merge them with images from the game world. These current and emerging MR techniques gain structure from game conventions. We are helped to believe in the world by the very focused actions and goals of the game.

Theme Park Rides are perhaps the most developed forms of interactive virtual reality with strong narrative immersion. The Disney design team has developed clear guidelines for the successful creation of these elaborate fantasy experiences. As I have noted elsewhere, the concept of a "ride" already sets up a useful immersive boundary [8]. Disney rides are 4-5 minutes long and involve physically boarding a vehicle of some kind. The rides rely on real world objects for interaction with virtual environment -- like swords and steering wheels -- so the designers do not have to teach interactors, or "guests" as they consider them, a new controller. The physicality of these objects make them ideal transitional forms: they can be felt in the real world but their actions have consequences in the virtual world. The rides are also set up so that the interactor has to make conscious choices of where to steer or which weapon to shoot. For example, on a ship ride, the guns were placed on either side so that kids would be encouraged to run back and forth shooting while parents sat in one place and navigated. The space is arranged to reflect story possibilities, with large structures attracting guests to steer toward story-rich areas. But the designers did not leave excitement to chance. If the guests steer their imaginary ship away from the exciting places, a storm will arise and blow their ship in the direction of dramatic events. The Disney rides carefully structure the guests experience, starting with the pre-ride wait which is used to show them related story elements and demonstrate gameplay, the first part of the ride which is meant to get them used to the game actions, the exploratory choices, the climactic battle, and the final theatrical sequence. The technology is only interesting to the designers insofar as it augments the experience of the interactor [5]. This is the opposite of the situation of most research labs where the technological development drives the process and the applications are secondary.

Still, research labs can learn from the Disney team **to exploit familiar story patterns and to use familiar physical objects to anchor the action of the interactor.** Research labs often ignore the obvious scenarios. For example, a recent augmented reality project at Georgia Tech involved catching butterflies with the use of a tube that rotated the sphere of butterflies to bring them closer. The tube was meant to address issues of tricky depth perception in augmented reality which makes it hard to tell how far away something is. Design critique suggested that a better user experience would be to make the tube into a net for catching the butterflies. The difficulty of depth perception forms a good research focus, but if the objective is to provide a satisfying user experience one might have some kind of butterfly bait in the net and shape the gameplay around inducing the butterflies to fly into the net, an idea suggested by a successful Eye-Toy scenario. Simplifying tasks so that it is easy to understand results is a crucial part of immersive design, as is eliminating confusions with equipment.

Brown and Cairns reinforce this observation, pointing out that transparent interfaces are the minimal requirement for an immersive experience and that any glitch in the action of the equipment or in understanding the gameplay disrupts engagement, the most basic stage of immersive experience.

3 Immersive Space and Point of View

One of the primary challenges of virtual worlds is establishing the **space** so that it is free of distractions, open to dynamic change, and dramatically compelling. Spatial presentation is a form of representational abstraction. In a narrative environment the presentation of space and place sets up dramatic expectations and reinforces dramatic themes. Pacman, SuperMario, and Doom rely on highly abstract physical worlds whose features have been adapted to the game tasks. Other games use narrative abstractions from fantasy literature or film noir to set up the action. The landscape and speed of navigation through Myst or Grand Theft Auto set up very different but equally rich narrative expectations. Augmented Reality environments will work best when they exploit specific locations such as a cemetery in Atlanta filled with graves of interesting historical figures (J. Bolter and B. MacIntyre, Georgia Tech, see http://www.lcc.gatech.edu/~bolter/oakland.htm).

Conventions of presentation can help to reinforce the reality. For example, the use of windows and doors as portals to dynamic or pre-recorded scenes or characters, is a convention that fits both the technical limitations of VR worlds and the real world expectations of the interactor. It also adapts a useful stage and film convention for allowing entrances and exits from a fixed part of the stage, carrying the suggestion of a larger world beyond what is immediately visible.

The work of the ICT draws upon principles of theatrical design and film set design. Their Flatworld project (viewable at http://www.ict.usc.edu/) starts not with the idea of reproducing the actual world, but with the abstraction of a "flat" – a room built for narrative purposes with windows placed on two adjoining walls so that a helicopter may be viewing in first one and then the other, reinforcing its presence and motion. Movies with live actors can now be produced using completely digital sets, but the sets are added in time-consuming post-production. In the pre-digital era they used a form of rear projection, with actors standing in front of still or moving images of Paris or the Sahara, or driving a car against a background of outdoor landscape. Similarly canned images, or dynamically created images from a set of highly recognizable objects could be used in a similar manner to deepen the illusion and set the scene. Providing the interactor with a flashlight in a simulated alley could greatly increase the sense of being there because the interactor's own actions would reveal clues to the murder or hiding assassins or betraying lovers or conspiring spies – or whatever other story element might have dramatic power within the specific fictional world. The flashlight would look and feel like an actual one but would work upon the virtual world.

Spacial presentation is particularly tricky in public spaces where there is increasing interest in Mixed Reality experiences. Since, as Brown and Cairns point out, distraction is is the enemy of immersion, designers of location-based MR face the problem of keeping interactors engaged though part of their attention is necessarily engaged in negotiating the real world, walking across a square, avoiding other pedestrians, perhaps even crossing streets. Reid et al who designed such an installation with position-specific audio recreating the sounds of an historical riot, suggest that people actually go in and out of immersion in this situation. They suggest that stronger narrative structure can keep interactors engaged with the virtual world for longer periods and help them reenter when attention is diverted to the actual world. They also suggest that for at least two of their 700 subjects the experience of audio-based Mixed Reality is one of "parallel worlds" [9]. In a visual mixture it may be harder to separate the experiences into coordinated but separate parallel experiences. And if narrative immersion is the goal, then it will probably be necessary to remove the experience to a specially prepared environment, such as a visit to a castle. It will also help if the appearances are virtual because they belong to the past and are therefore tantalizingly untouchable. Ghosts in a graveyard or a castle would be perhaps the ideal dramatic presence for this form. They could provide the kind of snapshot of the past and reflection upon mortality itself found in Edgar Lee Masters' *Spoon River Anthology* (1915);

or the might induce the kind of aching drama created by the scene in Thornton Wilder's play *Our Town* (1938) where a girl tries to reach across the barrier between the living and the dead to touch her mother. One can imagine a highly popular MR tour of Princess Diana's childhood home or a more didactic one of Thomas Jefferson's house. But the most involving narrative experience in Mixed Reality would probably be more like a ghost story, a horror movie, or a scare ride in an amusement part with the video ghosts played up for maximum eeriness, surprise, and gore. Because the conventions of this narrative form are closest to the affordances of the medium – the spectral presence of superimposed figures, the possibilities for appearing and disappearing in disconcerting ways, the separation between the actual people and the virtual artifacts which reminds us of a hallucination or a haunting – it is easier to see how to integrate the story structure and the interaction. For other forms of narrative, such as the ironic juxtapositions of Masters' monologists or the elegiac longing of Wilder's play we will have to work harder to find appropriate patterns of presentation and interaction.

Point of view is one of the most challenging areas of virtual worlds. Older genres like novels and movies have established conventions for indicating to the reader/viewer what they are permitted to see and whose perspective they are seeing it from. Games have multiple conventions including third person and first person, with and without a separate omniscient "camera" view. The Eye-Toy uses the Magic Mirror effect, capturing the image of the player, which is more appropriate to skill games than to imaginary worlds. The only such applications that I have personally found immersive are the ones that allow you to play with your visual image, as in the electronic funhouse mirrors of the San Francisco Exploratorium, or the award-winning installation art *Text Rain* which captures a shadow image of the user stretching out arms and hands to manipulate words falling down the screen. In *Text-Rain* the tumbling of the words over the figure, their responsiveness to bouncing motions, and their serendipitous arrangements create an abstract playground, but it is not a virtual place separate from the space in which you are standing. Magic mirrors work best, then, when they are used as mirrors not as Alice in Wonderland style looking glasses into another reality [6].

Although Marie-Laure Ryan thinks that placing the interactor in the virtual world makes it more immersive, it also makes it harder to maintain the immersion because our own bodies offer so many cues that we are not there. Placing the action at some distance from us through windows, doors, imaginary "alternate dimensions" as in ghostly appearances makes it easier to sustain the illusion. Vehicles like flight simulators or the rides in Disney World are ideal for giving us presence at a distance. Point of view is easiest to arrange when the interactor is only a navigator, flying around a world and viewing it without any representation of the person within the world.

Threshold objects can be useful in positioning the interactor as well as guiding their actions. Vehicles and concrete objects are excellent threshold objects, taking us not just into the illusion-making technology, but into the imagery of the imaginary world. A pirate ship or a canon is a better threshold object than a VR headset because it works in both the tangible and the illusory reality. But threshold objects can also be household items if they are positioned appropriately in the virtual world. In an augmented reality installation based on Alice in Wonderland, you may sit as an interactor at a real table with a real teacup in front of you. Putting on the headset, you see Alice, the Rabbit, and the Mad Hatter also seated around the table. If you move your (empty) teacup in a gesture of throwing tea at any one of them they get wet and react with comic anger. We can imagine installations like this developing into an extension of role playing games in which various table objects: letters, money, treaties, food form part of a shared reality for multiple distanced players joined by some virtual players. A key part of making such environments satisfying will be maintaining a consistent point of view as is done in films and narrative, and limiting what the interactor can see, hear, and touch appropriately so that it conforms to their physical position. It will also be important to distinguish between gestures that result in actions within the fictional world and actions that act on the representational apparatus. For example, changes in a camera angles can makes sense in a screen-based game but would be profoundly disorienting in a virtual reality unless there were some physically analogous user action, such using as a magnifying glass or pulling back in a helicopter or perhaps taking a reading from a wearable screen that provides a view from a virtual overhead camera. To maintain immersion the user must be anchored in a coherent physical space.

4 Interactions with Virtual Characters

Immersion is easiest to sustain in depopulated worlds like *Myst,* but narrative requires the presence of virtual characters. Gestural exchange like shooting or throwing tea is expressive but limited. Menu-based exchange is more expandable but the menu exists on a meta-level that can threaten the immersion. Conversation with virtual characters is a compelling goal of interactive storytelling, and in some ways it seems easy to achieve. The era of computational storytelling begins with the surprising success of Joseph Weizenbaum's Eliza who deceived users into thinking she was a real person though she had only simple pattern recognition and Weizenbaum's genius for scenario making [13] [14]. Later chatbots have uniformly failed the Turing test, but succeeded in fooling people in online chat rooms. Students using the CharacterMaker program I developed at MIT and am now using at Georgia Tech find it possible to create comically predictable characters who can sustain consistent conversations for as long as thirty exchanges.

Characters who talk to us achieve a high degree of presence in games, particularly when they draw on genre fiction conventions of villainy and when they show flexible autonomous action and independent goals within the game world. The characters in the *Last Express* are cinematic in many ways and gain reality because we can intersect with and eavesdrop on them carrying on their lives by moving around the train, meeting with one another according to their own schedules and desires, without regard to the protagonist's actions. The encyclopedic detail with which they go about their goals reinforces immersion, and prevents us from being frustrated by the fact that we cannot talk back to them except by menu choices. There are some reliable techniques for creating these characters so that the conversation is more likely to make sense. It is important to have the character feed the interactor key words by referring to things that can be echoed back. This allows the author to anticipate keywords and prepare appropriate responses. It is also important to have a varied supply of default responses that fit the character and the situation [8].

The most ambitious attempt to create characters who can carry on a spontaneous conversation with an interactor is *Façade* (2005) by Michael Mateas and Andrew Stern. In this story game the interactor is in the role of an old friend visiting the home of a quarreling young couple. The aesthetic of the game emphasizes the spontaneity of the interactor, encouraging people to use their own names to play and allowing them to interrupt the dynamically generated animated characters at any point. The interactor types in their contribution to the conversation while the computer controlled characters, Grace and Trip, speak in pre-recorded sound bites. The conversation gains continuity from the fact that Grace and Trip will continue to talk to one another if there is no input and they will also ask direct questions of the visitor. The utterances of the interactor are not understood literally but interpreted according to a schema of speech acts including agreeing or disagreeing with the husband or wife, praising or criticizing them, and using likely key words like “divorce” or “depressed”.

Façade pushes natural language as far as it can currently go, sometimes at the expense of immersion and story pleasure. The interruption mechanism, for example, does not work well with typing because the characters can’t tell when you have begun to say something but must wait for you to hit “enter”. Often they begin to say something interesting only to abruptly stop when the interactor “speaks.” There is no way to tell them to go back to what they were saying. This violates the turn taking and repair mechanisms that are important parts of natural conversation. The interaction is therefore a special kind of computer-human conversation, but it has not been optimized for that purpose.

In general allowing natural language input raises the level of expectation and inevitably creates frustration when utterances are not understood. In an MIT project of the 1980s, we tried to create a conversational game for language learning using language parsing and generation. The premise was a character who had lost him memory, which we thought would provide good defaults that were motivated by the situation. He could be confused or even break off the conversation if he did not understand the questions. We also thought that we had a good chance of understanding the interactor because as language students they would have limited vocabularies and we would know all the words they knew because we knew their textbooks. This turned out not to be the case. Users coined words or used words incorrectly or in unpredicted syntax. They brought all their knowledge of their first language or of other foreign languages to bear upon their constructions. We were able to create a system that had a coherent idea of a conversation and could generate good responses to well-formed input, but it could not deal with the variety of utterances that is the defining affordance of language as a system of communication, even among novice language learners [7].

As it is, there is a wonderful pleasure when the computer-based characters understand what is said to them. But it is not realistic to expect users to find sustained immersion in a world in which there is constant misunderstanding. For this kind of interaction to be successful users have to become used to the level of understanding that they can expect and then play to that. What is needed is a set of conventions around which we can plan such conversations and train users to participate in them. The gestures of the Sims are one source of non-verbal conventions and Façade makes use of them to some extent, allowing the interactor to “comfort” or “hug” the characters. Perhaps greeting, name-calling, praising, echoing words as questions, and other verbal rituals can be developed into a similar set of established conventions for conversations with computer-based characters.

It might make sense to put aside linguistic interactions in virtual worlds since the visual sense of presence would compound the effect of raising expectations through the use of language. Instead, we might try to exploit the single most powerful and basic act of human communication, pointing and gaze-following. Cynthia Breazeall’s [3] group at the MIT Media Lab is making impressive use of the gaze, nodding, and pointing to create scenarios in which users are set to teaching an appealingly babyish robot to do simple tasks.

The work of this group recreates the situation which may be at the root of the evolution of human consciousness and the invention of human game-playing: the shared attentional scene [12]. Cognitive scientists believe that human beings differ from our ape cousins in one fundamental evolutionary change that may underlie all of our cognitive advancements: the ability to imagine the consciousness of another person and to share attention with them. Story scenarios in virtual worlds might be based on mute or alien characters, or characters who speak a foreign language: a world without babelfish of any kind. Babies and parents communicate in this way before the emergence of speech. They are able to make themselves understood because of shared attention on objects of interest and common tasks such as feeding, playing peek-a-boo, shaking rattles, etc. The challenge for designers would be to invent a reality in which story progresses through a similar repertoire of shared gestures, with perhaps some non conversational language here and there.

The objects in the Sims are good examples of the use of pointing for narrative purposes, in part because they all contain menus of possible actions that can be performed with them. The menu is perhaps intrusive, disrupting the imaginary world with meta-information, although it becomes transparent with use. It is more dramatically satisfying to have objects with intrinsic plot possibilities that can be activated just by standing in front of them, grabbing them, or perhaps through words. For example, in Façade, the apartment of the quarreling couple is filled with objects that provoke revelations, including their wedding picture, a photograph of Italy, a painting, a bar with choices of beverages. The space is arranged as a kind of memory theater or props for improvised segments of plot. The visitor can trigger plot developments by moving closer to one character or another or by walking out the door.

5 Establishing Boundary Conventions

Digital environments are testing the possibilities of immersion through mixtures of the real and the virtual without recourse to VR technologies, but only using the internet, wireless devices, and other communication modalities. The internet puzzle gamethat was aimed at promoting *AI,* the Spielberg film, led to obsessive collaborative teams breaking the code of phony websites and related hoaxes that fed a conspiratorial frame of mind. The ill-fated ambitious EA game *Majestic*, released in 2001, happened in real time at unpredictable intervals, breaking into actual world through messaging devices. It challenged the fourth wall conventions in many ways, mixing its transmissions with participants email and faxes as well as embedding them on websites. But its most challenging transgression, perhaps, was its refusal of an appropriate temporal boundary. As Wired magazine reported:

> The game's incredible creep-you-out factor lies in its pacing -- and in your lack of control. If a character tells you he'll call you tomorrow, he's not kidding. You'll have to wait a day to get his information. Faxes may arrive without warning, and instant messages may pop up onscreen when you least expect them -- maybe even while you're at work. You might answer your phone only to be greeted by a frantic woman whose call is suddenly disconnected. Unlike every other game you can think of, Majestic doesn't wait for you to decide when it's time to play [4].

Majectic was often compared to David Fincher's film, *The Game* (1997), which captures the terror of participating in a make-believe that may or may not be real. We enjoy threat and violence as dramatic spectacle but we do not enjoy the actual experience. Popular entertainment often plays with this contradiction by creating ever more sensorily immersive experiences and by challenging the boundaries of make-believe. But Majestic crossed the boundary in an unacceptable manner by peddling imaginary terror in the fall of 2001 when Americans were gripped by the actual terror attacks on New York and Washington. The game had to be shut down because it was no longer entertaining to imagine oneself the target of unseen conspirators who can reach you where you live or work at any hour of the day.

Other art world games use internet, wireless, and geographical positioning technologies to purposely play with social boundaries. The Blast Theory group, for example, creates games that mix online players with street players and map a virtual city that corresponds to a real city, and a real place like an office that must be found by following a clue trail, to a virtual character, and to real world surveillance by web cam. Their Uncle Roy game explicitly asks the question "When can you begin to trust a stranger?" by offering street players limousine rides with unknown people and asking people to make one year commitments to unknown fellow players to help one another in an emergency. (http://www.blasttheory.co.uk/) . Although games like these seem far from the helmets or caves of VR technologies, they are similar in the discomfort they induce in interactors by disrupting social rituals and blurring the boundaries between the real and the virtual.

Immersion is easily disrupted on all of these dimensions: space, character, time, social conventions. When it works the experience is seamless. It begins with rituals of entering a separate space and assuming equipment that serves as a threshold into the experience. In a strong narrative environment the threshold should bear some relationship to the story world, such as entering a forbidden place or a magical vehicle or a particular stage set. The space seems present, detailed, and navigable. It provides not a reproduction of the real world but a carefully abstracted representation of a story world in which everything that is seen creates a consistent physical reality and creates expectations of story events. Moving around in this world is simple and consistent and the detail with which it is created reinforces the sense of its presence, allowing for the active creation of belief. The objects in the space are familiar or easily understood and they carry with them story possibilities, like a canon on a pirate ship or a wedding picture on the wall of a quarreling couple. User action is elicited and is always rewarded with an immediate effect that is consistent with and progresses the story. When user action is ambiguous, the story continues with default responses that also include cues to less ambiguous responses. As we get closer to achieving these design goals we are elaborating a concrete poetics of virtual reality.

References

1. Breznican, A.: Spielberg, Zemekis say video games, film could become one. Associated Press, San Diego Union-Tribune. San Diego (2004).
2. Brown, E., P. Cairns: A grounded investigation of game immersion. Conference on Human Factors in Computing Systems, Vienna, Austria, ACM Press (2004).

3. Breazeal, C., Gray, J., Hoffman, G., Kidd, C., Lee, H., Lieberman, J., Lockerd, A., Chlongo, D.: Tutelage and Collaboration for Humanoid Robots. International Journal of Humanoid Robotics I(2): 314-348 (2004).
4. McCauley, D.: Majestic Invades your World. Wired News (2001).
5. Mine, M.: Towards virtual reality for the masses: 10 years of research at Disney's VR studio. Proceedings of the Workshop on Virtual Environments 2003, Zurich, Switzerland (2003).
6. Moreno, E., B. MacIntyre, et al.: Alice's Adventures in New Media: An Exploration of Interactive Narratives in Augmented Reality. CAST, Bonn, Germany (2001).
7. Murray, J. H.: Lessons Learned from the Athena Language Learning Project: Using Natural Language Processing, Graphics, Speech Processing, an Interactive Video for Communication- based Language Learning. Intelligent Language Tutors: Balancing Theory and Technology. M. Holland. Mahwah NJ, Lawrence Erlbaum Associates (1995).
8. Murray, J. H.: Hamlet on the Holodeck: The Future of Narrative in Cyberspace. New York, Simon & Schuster/Free Press (1997).
9. Reid, J., Geelhoed, E. et al.: Parallel worlds: immersion in location-based experiences. Conference on Human Factors in Computing Systems, Portland, Oregon (2005).
10. Ryan, M.-L.: "Immersion vs. Interactivity: Virtual Reality and Literary Theory." Post Modern Culture 5(1) (1994).
11. Ryan, M.-L. Narrative as Virtual Reality. Baltimore, Johns Hopkins Press (2001).
12. Tomasello, M.: The Cultural Origins of Human Cognition. Cambridge MA, Harvard UP (2001).
13. Weizenbaum, J.: "Eliza-- A Computer Program for the Study of Natural Language Communication between Man and Machine." Communications of the Association for Computing Machinery Volume 9, Number 1 (January 1966): 36-35.(1): 36 (1966).
14. Weizenbaum, J.: Computer Power and Human reason. New York, Freeman (1976).

Formal Encoding of Drama Ontology

Rossana Damiano, Vincenzo Lombardo, and Antonio Pizzo

Dipartimento di Informatica and CIRMA, Cso Svizzera 185 Torino, Italy
Dipartimento di Discipline Artistiche, della Musica e dell'Arte and CIRMA,
Via S. Ottavio 9 Torino, Italy

Abstract. The goal of this research is to lay the foundations for a formal theory of drama, that abstracts from the procedural and interactive aspects involved in the generation of dramatic content. Based on the structural accounts provided from traditional drama analysis, the theory proposed in this paper exploits an agent-based perspective on characters to provide a goal-based characterization of dramatic qualities.

1 Motivations

Since the advent of digital media, character-based, narrative forms of communication have become commonplace in human-computer interaction, including user interfaces, entertainment, and education [1,2,3,4,5]. The need for autonomous behavior required by these applications has led scholars to adopt the agent techniques developed in artificial intelligence research to define and implement the virtual characters [6]. As applications have evolved into multi-character, interactive systems, the need for a centralized management of the plot execution has been realized by AI techniques for multi-agent coordination and cooperation [7,8].

While the main effort of system developers has addressed the use of AI techniques in the production of interactive storytelling applications, designers have relied upon the widely acknowledged corpus of drama studies - from Aristotle's investigations to structuralist approaches - to characterize the dramatic qualities of virtual narratives. However, there is still lack for computational theory that exploits the conceptual tools of AI to characterize the principles of drama. The aim of this paper is to lay the foundations of a formal theory that systematizes the basic aspects of drama in a direct and explicit model, with an immediate integration with agent-based theories. The theory, called *Drammar*, abstracts from the procedural aspects of drama generation, and is intended as the starting point for specifying, implementing and evaluating practical storytelling systems in a principled way.

Drammar is structured into two levels (see Figure 1). The *actional* level models the intentional behaviour of the characters in a plot by enforcing a BDI perspective on characters as intelligent, goal-directed agents: following Bratman's theory of practical reasoning [9], belief-desire-intention (BDI) agents form goals to pursue their desires, and, given their beliefs about the world, devise plans to achieve them. This level is augmented with a representation of emotions as provided by the OCC model [10]. The *directional* level accounts for the realization of a direction through the plot, by abstracting from the intentionality of the characters through the use of attributes that model the effect of plot incidents onto the characters' (i.e. agents') mental and emotional state.

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 95–104, 2005.

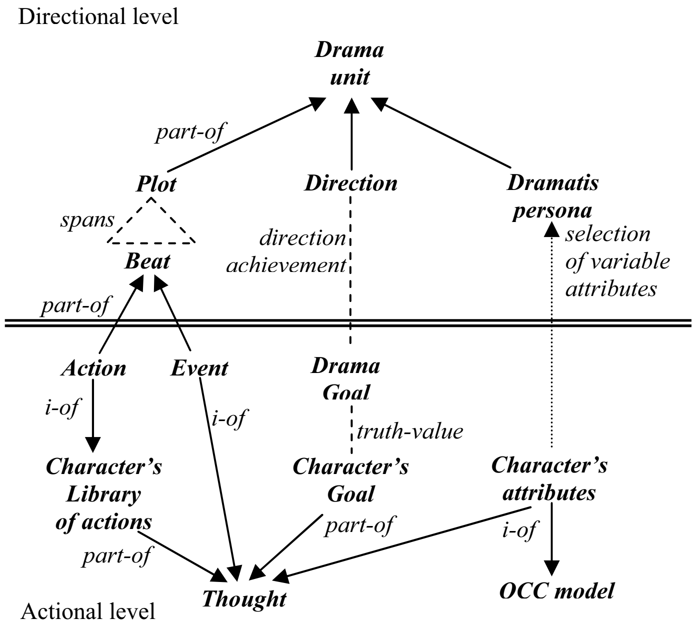


Fig. 1. Drammar ontology. The higher level represents the directional level, the lowest level represents the anchoring of directions in actions, i.e. the actional level.

In this paper we motivate and introduce the directional level of Drammar and we apply it to a representative example of dramatic art, a fragment of Hitchcock's film *North by Northwest*. The paper first introduces the relevant issues, from the literature on drama analysis, that are to be encoded in Drammar, and then presents a formal definition of Drammar. An example and the conclusions close the paper.

2 Drama Analysis

The goal of a drama is to make audience perceive what is intuitively called a "story" by exhibiting the actions of some *characters* (in conflict); actions are organized in a *plot*; the plot moves toward a *direction*. The notions of direction, character (also called *dramatis persona*) and plot, pervasive throughout the literature on drama analysis, are the three components of the directional level of Drammar (see figure 1).

The drama direction derives from the notion of "unity of action", originally expressed by Aristotle [11] and clearly stated by Stanislawsky and Styan [12,13]. An operational account of direction is the notion of *premise*, which takes a tripartite form ⟨*Character's value* at stake, *Conflict* (or direction of action), *Results* (or the end)⟩ [14,15]. The premise contains a movement caused by a conflict between one character and the environment and/or other characters. Such a movement concerns a rational or emotional value of the character. As we will illustrate, in Drammar, the *drama direction* is a change function that transforms the initial state of a character into a different final state at the end of the drama [14,16]. Such states are defined through values assigned to a number of *attributes*, or dramatic qualities of the characters. For example, given a premise according to which “greed leads to ruin”, the *self-satisfaction* attribute of a greedy character may change its value through a sequence of actions.

The relationship between the value changes of the rational/emotional states of the characters and the actual actions and events listed in a drama is stated through the notion of *drama goal*. The drama goal is the world state brought about by the plot, that realizes the drama direction. In particular, the drama goal is operatively specified as the achievement or the negation of the goal of a character, the drama *protagonist*. In the example above, the drama direction consists in the greedy character loosing his self and social esteem; an appropriate drama goal may be the state in which the greedy character is caught in the act of stealing the alms. The other characters, by acting for the achievement of their own goals, act as to oppose the protagonist (the *opponents*), or to help him (the *extras*). The combination of the actions acted by the protagonist and the other characters, together with the environment events, achieve the drama goal. For example, the goal of the greedy character to steal the alms without being caught will be frustrated by the an opposite world state provided by the drama goal: as a consequence, the stealing plan of the greedy character will be opposed by other characters, who may catch him red-handed.

The drama direction is defined as a rational/emotional state change over a set of attributes that define the character, or *dramatis persona*. This set is divided into two sub-sets, in order to combine the rational, Belief-Desire-Intention (BDI) perspective [9] on agents with an emotional component. The rational attributes model the character’s knowledge, subdivided into ontological, actional, social and normative, and her/his motivation, or desires (the agent’s long-term goals and preferences), that will translate into the intentions pursued by the agent at the actional level.[1] The emotional component is structured along the lines provided by the Ortony-Clore-Collins (OCC) cognitive model of emotions [10], widely employed in interactive drama systems [18,19,8]. In this model, emotions basically derive from the appraisal of events, action, and objects in terms, respectively, of their desirability for the agent’s goals, their praiseworthiness according to the agent’s moral standards, and their appealing according to the agent’s dispositional attitudes. Attributes are valued positively or negatively; a value change of an attribute can go from positive to negative or vice versa. In the example of the greedy character introduced above, the drama direction is accomplished by the change of an emotional attribute, *self-satisfaction*, which models an agent’s approving (or disapproving) of the consequences of her/his own actions; this attribute turns from positive to negative as a consequence of the failure of the character’s goal of stealing the alms.

Concerning the plot structure, it has been a well known convention, dating back to Horace’s Ars Poetica [20], to segment the list of actions that form a drama into a number of units or sections, be it a “sequence of impressions” with a non-specified length [13] or a more clear-cut subdivision [21,22,23,24]. Segmentation has been formalized in semiotic studies in order to mark the discrete advancements of the narration [25,26]. It is important to notice that such units, despite terminological disparities, are of the same nature, so that some authors define drama as a recursive or “fractal” structure [23]. We

[1] In drama writing techniques, a similar account of characters as entities articulated into significant features, classified along personality dimensions, is put forth by Seger [17].

call the units of this recursive structure *drama-units*, and define drama as a drama-unit that is not included in any other unit. In parallel with the inclusion of drama-units in larger drama-units, the same relation holds between drama goals: in the example of the greedy character, the drama-unit characterized by the drama goal of stealing the money contains a smaller drama-unit characterized by the goal of finding where the money is kept. Following McKee [24] and Elam [26], at the very end of the recursive structure there are the minimal units of the plot, called *beats*, pure actional units formed by a action-reaction pair. The actual changes of the attribute values occur in beats, and every beat is part of a least one drama-unit. Going back to the example of the greedy character, the failed stealing of the alms may be the last unit in a sequence through which the character comes to know about the place where the alms have been hidden (first unit), then conceives the idea of stealing them (second unit), and eventually goes into action (third unit). In the same way as the emotional change in the character's self-satisfaction was modelled by the change of the *self-satisfaction* attribute, the changes brought about in the character by the first two units can be modelled through specific attributes that model the rational connotation of the character: the fact that the greedy character comes to know about the hiding of the alms (a belief) is modelled by the change from negative to positive of the *knows_about_hiding* attribute, and the fact that he conceives the idea of stealing them (an intention) is modelled by the *want_steal_alms* attribute (accompanied by an emotion of *hope*), which takes a positive value.

The description so far implicitly assumes the we are working with linear drama, where a unique list of beats is determined by a unique hierarchy of drama-units. It is worth pointing out here that we do not claim that each drama has a unique interpretation of its plot structure in terms of beats and drama-units, since each individual in the audience may perceive a different structure, but that each interpretation corresponds to a unique plot structure. So, each interpretation proposed by the literary criticism about a specific drama maps to a unique plot structure. In non-linear drama, there is not a unique list of beats, but a multiplicity of plots, licensed by a formal system (e.g., a formal grammar, a constraint-based system, etc.). Non-linear drama generalizes the case of linear drama by resorting to a meta-system to generate the plot structure: the possible lists of beats licensed by the meta-system are alternative realizations of the drama, and a linear structure is a special case that results from a non-ambiguous meta-system. In an interactive version of the greedy character story, the user may decide whether the character is to be caught red-hands or not: keeping the drama direction unchanged, if the character succeeds in stealing the alms, he may for example contract a mortal illness because the alms were infected. Or else, the top level of the plot structure itself may change to a different one, for example, "ruin leads to redemption", and the character, once in jail, may redeem himself. Notice that the set of the characters' variable attributes associated with the drama-units can be contradictory, since the multiplicity of plots can realize multiple drama directions and different characters' sets. It is a matter of future research to determined what elements should be kept fixed (or unique) in order to achieve the coherence of the plot.

The units of the drama and their directions are combined in a drama-specific progression that is related (via the protagonist's fate) with the emotional engagement of the audience. Dramatic actions in the plot trace a curve related to the fulfilment of the direction. Each drama-unit, with its goal, has both a temporal position and a *dramatic value* within the plot. The dramatic value of a drama unit is given by the number of value changes that occur within the unit, either in a beat directly included by the unit or in a beat included in some sub-unit of the unit itself. The temporal position of each unit is mapped onto a dramatic value based on the content of the unit, and the resulting succession of values forms a curve, the *dramatic arc*. The drama literature widely acknowledges an upside down U-shaped form of the dramatic arc: the first, the highest and the last dramatic values have received specific names (inciting incident, exposition, or rising conflict; climax, crisis, or turning point; conclusion or resolution [14,21,22,23,24]. If a drama is carefully designed, the dramatic arc reaches its peak towards the end of the plot (some changes may occur later).

3 The Drammar Ontology

The ontology of drama presented here, called Drammar, consists of two levels, a *directional level* that encodes the specific traits of drama illustrated in the previous section and an *actional level* that connects such traits with an agent-based perspective on characters. At the directional level, a Drama-unit is a triple of Plot, Direction and Dramatis Personae. The actional level unfolds the rational and the emotional perspective in terms of the facts that occur in the Drama-unit, the goal of the Drama-unit, and the attributes of the characters. The facts can be either *Actions*, i.e. facts determined by the characters in pursuit of their aims (*Characters' Goals*) and belonging to the *Characters' Library of actions*, or incidental *Events*. The beliefs and long-term desires as well as the emotions of the characters are encoded in the *Characters' attributes*, which in fact are instances of the rational knowledge (Thought in Aristotle's terms) or of the emotion types contained in the OCC model respectively.

Each element of the directional level is connected with some element at the actional level through some specific relation. The actions and events represented in the drama at the actional level are listed in the Plot only as functional to the accomplishment of the Direction. The Direction, in turn, models the goals of the characters according to a fulfilment/frustration dimension, through which the characters' dramatic qualities - a selection of their overall specification - are affected. The directional level of Drammar is encoded in the notion of Drama-unit, which consists of a number of Dramatis personae, a Direction and a Plot (see Figure 1). Drama-units are organized in a hierarchical structure, which accounts for the segmentation of drama. We now present each component; then we provide the definition of the overall notion of drama.

The Dramatis persona is a set of $\langle$attribute, value$\rangle$ pairs, where the attributes are binary-valued $\{$+,–$\}$. A drama-unit inverts the values of one or more attributes of some characters. Not all of the attributes describing a character at the actional level will be affected by the changes carried out by the drama plot. For example, Othello changes his attitude toward Desdemona but does not resign from his military position; Nora stops performing her stereotyped married life but does not stop loving her children.

Definition: A dramatis persona $CHAR$ is a pair $\langle ATT, POLARITY\rangle$, where ATT is a subset of $POOL$; $POLARITY$ is a set of pairs $\langle x, v_x\rangle$, where $x \in ATT$ e $v_x \in$

$\{+, -\}$. All the attributes in ATT are assigned a value in $POLARITY$ and for each attribute only one assignment is permitted.

The Direction is a function D that specifies the value changes of the characters' attributes after the execution of the Plot element of a Drama-unit. So, the domain of the direction function is a State (where a State is a set of Polarities of attributes), and the co-domain is another State. So, let a State be $\bigcup_i CHAR_i.POLARITY$:

$$\mathcal{D} : State_j \rightarrow State_f$$

where we enforce the *Minimal Direction* condition that at least one attribute inverts its polarity, i.e. there exists at least one attribute $a_i \in CHAR_i$ such that the value assignment of a in the initial state $State_j$ is different from the value assignment of a in the final state $State_f$.

The drama-plot is the component that carries out the polarity inversions as described by the Direction function. It consists of a list of actions/events grouped in Beats. The value changes required by the Direction function occur within one Beat. Notice that some Beat may not change any attribute value, but every change does occur in some Beat. The three components above form a drama-unit.

Definition: A drama-unit is a triple $\langle Dramatis\_personae, Direction, Plot \rangle$, where:

- $Dramatis\_personae$ is a finite set of Dramatis_persona $\{DP_1, DP_2, ..., DP_n\}$;
- $Direction$ is a function D defined as above;
- $Plot$ is a list of Beats $\langle B_1, B_2, \ldots, B_m \rangle$,

and the Minimal Direction condition holds.

A drama-unit represents by itself all the basic aspects of drama, as introduced in the previous section. Drama-units are subdivided into smaller drama-units, until the level of elementary drama-units is reached. The resulting structure is a tree of drama-units, whose leaves are directly connected to beats, and whose root is the properly called drama, the highest-level unit that subsumes the entire sequence of beats. Formally, a drama is the drama-unit that is not dominated by other drama-units.

The dramatic value of a drama-unit is provided by the number of value changes that occur within the unit or in some sub-unit. A drama-unit changes values either directly (that is, in a beat that it includes directly) or through a sub-unit (that is, in a beat included in a sub-unit). The horizontal (temporal) position of a drama-unit is the beat where the last value change performed directly by the drama-unit occurs. Given a drama-unit, the dramatic arc of that drama-unit is given by the line formed by connecting all the dramatic values of the children units plus the unit itself.

4 An Example

In this section, we apply the formal system Drammar to the well-known Hitchcock's North by Northwest [27], following the analysis reported in [23]. North by Northwest is about a middle-aged advertising executive, Roger Thornhill, who is mistaken for a government agent George Kaplan by a gang of spies lead by Mr Vandamm. He gets involved in a series of misadventures and is pursued across the States by both the spies and the government whilst being helped by a beautiful blonde Eve Kendall. Eventually he will discover that Eve is an undercover CIA agent and together they will defeat the evil gang, on a thrilling sequence on the Mount Rushmore.

ID	Description	Drama Goal	Attribute	Value	Attribute-type	Dramatic Value
1	R. mistaken for Kaplan and kidnapped by Vandamm's gang	Kidnapped (Roger) True	Distress	+	EMOTION.well-being	1
2	R. gets aware of mismatch and tries get out of trouble	Involved (Roger) True	Individualism	-	BELIEF.norms	20
2.1	R. meets Vandamm	Agreement (Roger,Vandamm) False	Disappointment	+	EMOTION.prospect-based	4
2.1.1	Vandamm addresses R. as Kaplan	Mentioned (Vandamm,Kaplan) True	Distress	+	EMOTION.well-being	1
2.1.2	Vandamm threatens R. of death	Threatened (Vandamm.Roger) True	Anger	+	EMOTION.well-being/attribution	1
2.1.3	Vandamm's gang tries to kill R.; R. escapes	Killed (gang, Roger) False	Relief	+	EMOTION.prospect-based	1
2.2	Nobody believes R.; R. accused of shooting Townsend	Outcast (Roger) True	Isolation	+	BELIEF.world-state	4
2.2.1	R.'s report not believed by anybody	Discredited (Roger) True	Anger	+	EMOTION.well-being/attribution	1
2.2.2	R. leaves his mother	Left (Roger.Mother) True	Submission	-	BELIEF.social	1
2.2.3	R. is believed to have killed Townsend	Falsely_accused (Roger.assassination) True	Disappointment	+	EMOTION.prospect-based	1
2.3	R. escapes police, meets Eve, seduction, fake appointment	Seduced (Eve.Roger) True	Love	+	EMOTION.attraction	6
2.3.1	R. runs away by train	Caught (Roger.Train) True	Relief	+	EMOTION.prospect-based	1
2.3.2	E. hides R. from police in the cabin	Hidden (Roger) True	Gratitude	+	EMOTION.well-being-attribution	1
2.3.3	R. and E. sleep together	Had_sex (Roger.Eve) True	Satisfaction	+	EMOTION.prospect-based	1
2.3.4	E. fixes the fake appointment with Kaplan	Deceived (Eve,Roger)True	Hope	+	EMOTION.prospect-based	1
2.3.5	Airplane tries to kill R.	Meeting (Roger,Kaplan) False	Disappointment	+	EMOTION.prospect-based	1
2.4	R. calls E.'s bluff and Professor explains	Explain (Professor,Roger) True	Anger	+	EMOTION.well-being-attribution	5
2.4.1	R. discloses E.	Deceive (Eve,Roger) False	Reproach	+	EMOTION.attribution	1
2.4.2	R. finds about Vandamm and E.	Unmasked (Roger.Vandamm) True	Anger	+	EMOTION.well-being-attribution	1
2.4.3	R. arrested and meets Prof.	Meeting (Professor,Roger) True	Truth	+	BELIEF.world-state	1
2.4.4	E.'s identity revealed	Revealed (Eve's identity,Roger) True	Pity	+	EMOTION.fortune-of-others	1
3	R. takes revenge	Married (Roger,Eve) True	Family	+	BELIEF.norms	8
3.1	E. pretends shooting R. at M. Rushmore	Collaboration (Roger,Eve) True	Relationship	+	BELIEF.social	3
3.1.1	E. fake-shoots R.	Deceived (Roger,Vandamm) True	Satisfaction	+	EMOTION.prospect-based	1
3.1.2	E. to leave with Vandamm	Coupled (Roger,Eve) True	Love	+	EMOTION.attraction	1
3.2	Chase and fight at M. Rushmore	Saved (Roger.Eve) True	Gratification	+	EMOTION.well-being/attribution	4
3.2.1	R. escapes from hospital	Rebellion (Roger,Professor) True	Independence	+	BELIEF.normative	1
3.2.2	Leonard discloses Eve's secret	Informs (Leonard,Vandamm, Eve's trick) True	Fear	+	EMOTION.prospect-based	1
3.2.3	R. kills Leonard on M. Rushmore	Killed (Roger,Leonard) True	Relief	+	EMOTION.prospect-based	1

Fig. 2. Analysis of North by Northwest

The table in Figure 2 contains the translation of North by Northwest in Drammar terms. The first column, ID, reports the hierarchical structure of Drama-units; the levels of the hierarchy correspond (in decreasing order) to acts, sequences and scenes in the standard filmic terminology. For example, Act 1 - being introductory - is not subdivided into smaller units, while Act 2 and 3 consist of 4 and 2 sequences respectively.

The second column, Description, contains an informal description of the Drama-unit. The third column, Drama goal, contains the drama goal through which the direction of the Drama-unit is accomplished. The state to be accomplished is described as the state in which a certain predicate is true or false. For example, Act 2 (ID 2), leads to a state in which the moral standards of the protagonist, Roger Thornill, have been affected, so as to make him more inclined to help the others (the predicate "Involved(Roger)" becomes true). This drama goal will be in turn accomplished through the drama goals of the sub-units that compose it: the accomplishment of this transformation will be carried out by Roger's need to take himself out of a big trouble (developed in Sequences 2.1 and 2.2), together with the seduction operated on him by Eve (Sequence 2.3) and the awareness of a conflict between the intelligence service of his country and a group of evil spies (Sequence 2.4).

The last three columns, Attribute, Value and Attribute-type, describe the direction of each Drama-unit. In Drammar, the Direction consists in the change of the value of one or more attributes. For example, in Act 1, Roger falls into distress as a consequence of being kidnapped by Vandamm's gang, setting the emotional attribute "distress" to a positive value; in Act 2, Roger's "individualism", initially set to "–", is set to a "+". Although the attributes of the characters affected by changes belong to both the rational and emotional category (individualism and distress are representative of the two, as reported in the Attribute-type column), most attributes belong to the emotional type, especially in lower-level drama-units. The subtype to which each attribute belongs is expressed by the dot notation: for example, the "BELIEF.norms" expression referred to "individualism" means that this attribute belongs to the normative component of the character's rational connotation. When emotional attributes occur, their classification refers to the OCC cognitive model of emotional appraisal. For space reasons, the relation between the emotion activation and the character's appraisal is not explicitly represented in this table; for each attribute, we only report to which class of emotions it belongs according to OCC classification of emotion types. For example, in Scene 1 of Sequence 4 of Act 2 (ID 2.4.1), Roger realizes that Eve, after seducing him, is acting against him: we classify his mental state as dominated by reproach, as an effect of the moral evaluation of the behavior of Eve, construed as an agent. However, it is important to notice that the same situation could be as well construed as an undesirable event by Roger, and thus be a source of distress. Finally, the undesirable effects of Eve's blameworthy behavior for Roger may determine Roger's anger.

The last column (Drama value) reports the dramatic value of each drama-unit. For space reasons, we do not report a diagram of the dramatic arc, since its U-shaped form can be appreciated only by considering the temporal position of the beats where the attribute changes occur.

5 Conclusions and Future Work

In this paper we have provided the foundations for devising a formal theory of drama analysis. Although a number of systems have been developed for the generation of dramatic content, spanning from explicitly drama-oriented systems [28,1] to narrative systems [29,8,30], a formal characterization of the output generated by these systems is still lacking. Based on a broad analysis of traditional literature on drama cast in AI terms, Drammar accounts for the definition of drama on two levels: a directional level, which specifies the qualifying features of drama in a functional way, and the actional level, which models the actions acted by the characters in the plot in an agent-based perspective.

The Drammar system provides a language for describing drama as an off-line object, independent of specific applications. It relates the realization of a drama direction to the achievement of a goal at the actional level, but allows the same direction to be realized by different goals, and the same goal to be brought about by different courses of action and events. Interactive drama takes advantage of the non-determinism in the accomplishment of the direction, by letting the user navigate the space of alternative plots. From the off-line, ontological point of view, it does not matter if the choices that determine the final plot are accomplished by the author before the user's fruition or by a collaborative process between the author and the user through a procedural system.

The description of drama ontology provided in this paper is only a first step toward a comprehensive formal system, that includes a decision procedure for analyzing or generating drama. Several instruments have been proposed to this aim, ranging from story-oriented approaches (rooted in story grammars [31] and scripts [32]) to character-based systems, in which the story emerges as a result of the interplay of the behavior of multiple agents. In this paper, we have specified the requirements of drama in a perspective of system design and evaluation, and leave to future research the task of identifying the instruments by which the formal theory may be incorporated into practical systems.

References

1. Bryan Loyall, A., Bates, J.: Personality-rich belivable agents that use language. In Lewis Johnson, W., ed.: Proc. of the First International Conference on Autonomous Agents. (1997)
2. Staller, A., Petta, P.: Towards a tractable appraisal-based architecture. In Canamero, D., Numaoka, C., Petta, P., eds.: Workshop: Grounding Emotions in Adaptive Systems. (1998)
3. Marsella, S., Johnson, W.L., LaBore, K.: Interactive pedagogical drama. In: Proc. of Agents 2000. (2000)
4. Cavazza, M., Charles, F., Mead, S.: Interacting with virtual characters in interactive storytelling. In: Proc. of the First Interantional Joint Conference on Autonomous Agents and Multiagent Systems. (2002)
5. Szilas, N.: Idtension: a narrative engine for interactive drama. In: Proc. 1st International Conference on Technologies for Interactive Digital Storytelling and Entertainment (TIDSE 2003), Darmstadt (Germany) (2003)
6. Cassel, J., Bickmore, T., Campbell, L., Vilhjamlmsson, H., Yan, H.: Human conversation as a system framework: Designing embodied converational agents. In Cassell, J., Sullivan, J., Prevost, S., Churchill, E., eds.: Embodied Conversational Agents. The MIT Press, Cambridge, Massachussets (2000) 29–63

7. Mateas, M., Stern, A.: Towards integrating plot and character for interactive drama. In: Working notes of the Social Intelligent Agents: The Human in the Loop. AAAI Fall Symposium Series (2000)
8. Theune, M., Faas, S., Heylen, D., Nijholt, A.: The virtual storyteller: Story creation by intelligent agents. In Goebel, S., Braun, N., Spierling, U., Dechau, J., Diener, H., eds.: Proc. of TIDSE 03 (2003)
9. Bratman, M.E.: Intention, Plans, and Practical Reason. Harvard University Press, Cambridge (MA) (1987)
10. A. Ortony, G.C., Collins, A.: The Cognitive Stucture of Emotions. Cambrigde University Press (1988)
11. Aristotle: On the Art of Poetry. I. Bywater (transl.) 1920. Clarendon, (Oxford)
12. Stanislawski, K.S.: An Actor Prepares. Eyre Methuen, London (1980 (1936))
13. Styan, J.L.: The Elements of Drama. University Press, Cambridge (1963(1962))
14. Egri, L.: The Art of Dramatic Writing. Simon and Schuster, New York (1960 (1946))
15. Esslin, M.: The Field of Drama. Methuen, London (1988 (1987))
16. Hamon, P.: Pour un statut sémiologique du personnage. In: Poétique du récit, Paris, Editions du Soleil (1977)
17. Seger, L.: Creating Unforgettable Characters. Henry Holt and Company, New York (1990)
18. Elliott, C.: The affective reasoner: A process model of emotions in a multi-agent system. Ph.d. thesis, Institute for the Learning Sciences, Northwestern University (1992)
19. Bates, J., Loyall, A., Reilly, W.: An architecture for action, emotion, and social behaviour. Artificial Social Systems, LNAI 830 (1994)
20. Horace: Epistles, Book II; and Epistle to the Pisones (Ars poetica), edited by N. Rudd. Cambridge University Press, Cambridge (1989)
21. Field, S.: The definitive guide to screen writing. Ebury Press, London (2003)
22. Hatcher, J.: The Art and Craft of Playwriting. Story Press, Cincinnati, Ohio (1996)
23. Lavandier, Y.: La dramaturgie. Le clown et l'enfant, Cergy (1994)
24. McKee, R.: Story. Harper Collins, New York (1997)
25. Greimas, A.J.: On meaning: selected writings in semiotic theory. Pinter, London (1970)
26. Elam, K.: The Semiotics of Theatre and Drama. Methuen, London and New York (1980)
27. E.Lehman: North by Northwest. Directed by A. Hitchcock. Photographed by R. Burks. With C. Grant, E. M. Saint, J. Mason. Metro Goldwyn Mayer (1959)
28. Mateas, M.: Interactive Drama, Art and Artificial Intelligence. Phd. dissertation, Carnegie Mellon University (2002)
29. Bringsjord, S., Ferrucci, D.: Inside the mind of Brutus, a storytelling machine. Lawrence Erlbaum (1999)
30. Riedl, M., Young, R.: An intent-driven planner for multi-agent story generation. In: Proc. of the 3rd Int. Conference on Autonomous Agents and Multi Agent Systems. (2004)
31. Rumelhart, D.: Notes on a schema for stories. In Bobrow, D., Collins, A., eds.: Representation and Understanding: Studies in cognitive science, New York, Academic Press (1998)
32. Schank, R.C., Abelson, R.P.: Scripts, Plans Goals and Understanding. Lawrence Erlbaum, Hillsdale. NJ (1977)

Emotional Spectrum Developed by Virtual Storytelling

Nelson Zagalo, Ana Torres, and Vasco Branco

Department of Communication and Art, University of Aveiro,
3810-193 Aveiro, Portugal
{ntz, atorres, vab}@ca.ua.pt

Abstract. We have performed a quantitative study in order to find out the emotional spectrum of Virtual Storytelling in comparison with movies, taking into consideration the Russell's emotional circumplex model. Via internet forums we gathered videogames that people considered to be capable of eliciting each of the seven basic emotions, distributed around the circumplex. From the 200 videogames collected, we chose 14, following the principle of the two most cited for each of the seven emotions. These videogames were then tested with 33 subjects. These results were compared with Gross & Levenson (1995) study on movies. We found that these videogames were capable of successfully eliciting emotions such as Surprise, Anger, Disgust and Fear. There is also evidence that Happiness could be elicited. It was not possible to verify the existence of Tranquility. The most problematic was Sadness, except when interactivity was absent and emotion propelled through cutscenes[1].

1 Introduction

Storytelling can be seen as the "art of developing coherent event structures in time in order to provoke a determined emotional response and a cognitive change in the audience" [18]. In other words storytelling can be described as *the art of generating and managing expectations.* It is our belief that to produce successful virtual stories we need to be able to produce story experiences that can affect a diverse emotional spectrum of responses in the user. This takes us to the core of this study. From a psychological point of view this study is intended to try to find out what kind of emotional responses games are nowadays already able to generate and what problems and assets we do have in terms of emotional generation when using this new media to tell stories. Thus, we have analyzed which the emotions and the emotional spectrum were elicited by presented games (using Russell Model, see Fig.1). The results were then compared directly with what Gross & Levenson [6] stated in terms of emotion elicitation using movie clips. This allows us to show what do we have accomplished already and what is missing in emotional capacities in virtual storytelling.

Also, we suppose that games aren't capable to elicit Sadness because sequences that are thought to elicit always use film mechanics (as cutscenes). And if we attend

[1] Cutscene – "sequence in a video game over which the player has no control" (http://en.wikipedia.org/wiki/Cutscene).

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 105 – 114, 2005.

to Lazzaro's [9] study we can also see that the emotions purposed did not include Sadness emotion or equivalent. The systematic introductions of cut scenes in games alert us to the difficulty to elicit Sadness interactively. Thus, the aim of this study is also to test the possibility of eliciting Sadness interactively.

2 Theoretical Foundations

Gross [6] tested a set of films, in order to know if they were effective elicitors of emotions. Rottenberg [14] affirms that Gross's study [6] appears as notable effort to assembling a standardized library of emotion stimuli capable of eliciting emotional states. As Gross [6] study is applicable to a videogames study we have decided to follow similar methodology in order to establish comparisons with their results and to take conclusions upon videogames' emotion elicitors capacities.

There are three emotions' components, which are Emotion Experience, Behaviour and Physiology, as Vaz Serra [17] reviewed. We can come across many authors who defend those emotion response components (e.g. [13], [12], [16], [11], [8], [5] and [7]). Nevertheless, we have decided to evaluate emotions only by assessing subjective experience like Gross [6] did, in spite of evaluating other emotional components due to the fact that there is empirical evidence that emotional response systems do not perfectly covary in their activity and even are dissociated [14].

Gross [6] study suggests that it is possible to elicit discrete emotions with film clips. Our specific purpose is not to test if videogames elicit discrete emotions but if they elicit a multiple spectrum of emotions at least as films clips can do, in terms of basic emotions - joy, sadness, fear, anger, surprise, disgust – [3], [4] and [15].

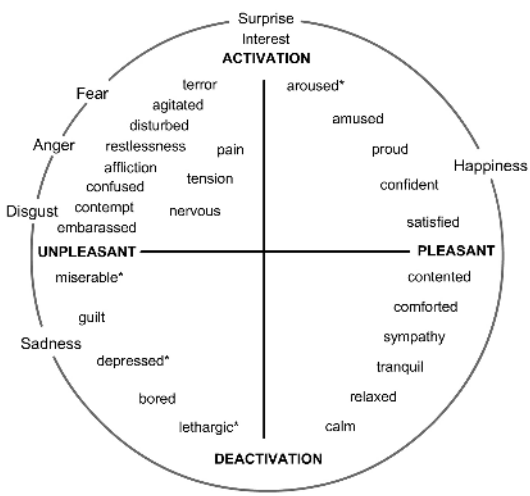


Fig. 1. Circumplex Model of Emotions [15]. This model contains already all the emotions from our questionnaires distributed among the quadrants. The emotions with (*) means they were not used in our questionnaire.

Russell developed the first “Circumplex Model of Emotions” in 1980. The purpose of this model depicts those emotions “can be described as a space formed by two bipolar, but independent dimensions, degree of Pleasantness and degree of Activation” [1]. In order to analyze the capacity to elicit a diverse emotional spectrum using games, we have considered the four quadrants as targets looking for our game excerpts, based in each of the basic emotions.

3 Study Methodology

A total of thirty three persons (27 men, 6 women) participated in video game-playing 90-120 minutes long individual sessions. All of them had played the selected 14 different games. The sample is heterogeneous in spite of Academic Degree (Undergraduate, Master Degree and Doctoral Student) and Employment (thirteen different jobs). But it is homogeneous in spite of residence area, because 69.7% are resident in the same city area (Aveiro) and 90.9% are resident in the same district area.

We requested volunteers, with one special pre-requisite which was the interest to play and enjoy the experience. It is interesting to verify that our sample (18-40 years old and more male than female subjects, 81.8% - 18.2%) has similar characteristics with target market of videogame industry. As pointed by the Microsoft's corporate vice-president J Allard, at E3 2005 “we're relying mainly on one type of consumer…the 18 to 34 year old male, he's the backbone of the industry" [10].

Physical Setting

The Physical setting in which games were presented followed the suggestions pointed by Rottenberg et al. [14]. The session’s occurred in a University laboratory in Aveiro, Games were restricted to PC platforms and Playstation 2 (PS2), in order to minimize physical interfaces learning.

There was always the presence of an experimenter to help with some doubts the subject could eventually have and to stop the game at the end of the selected sequence.

Pre-sessions Procedure

First of all, we solicited suggestions of games that would be effective elicitors of emotions (Happiness, Anger, Disgust, Fear, Sadness, Surprise and Relaxation), in order to use them in our experimental sessions. We asked people over Internet forums and we asked colleagues and game players through email and instant send messenger (9) too. We obtained 93 answers to the Internet forums that we have created, and also other forums over the Internet related to our subject (432 extractable answers) as well as an Internet pool on the subject of emotions (287). We gathered a total of 821 answers. At the end, from the 200 videogames referred we chose 14, the 2 most cited for each emotion.

In our games selection, we follow a simple rule to the emotional comparison between games and movies: game excerpts should permit any kind of interaction or representation manipulation in order to have any noticeable difference in terms of communication language when compared to movies.

Sessions Procedure

Firstly, subjects answered a 32 rating scale emotional pre-questionnaire[2]. The rating scale presented consists in a 9 points scale, in which the first one is the "0 = not at all", the middle is the "4 = some" and the last is the "8 = extremely". The aim of this self-report questionnaire was to evaluate the previous emotional level and then compare it to videogames post presentation level. After the game sequence subjects were requested to complete a similar questionnaire (Post-questionnaire).

We had used two different game presentation orders (Normal - 28 persons and Inverted - 5 persons), because we wanted to test the influence of this variable on final obtained emotional results. Firstly, we have always presented a game that we expected to elicit low levels of contentment (Max Payne - Normal or Final Fantasy VII - Inverted). It was never shown two games expected to target the same emotion in a row or more than three games of a particular valence consecutively. Normal order coincides with Appendix 1 games order and the Inverted consists on the inversion of each game with the immediately next to it.

At the end, we asked subjects to answer a Post Experimental Interview Questionnaire, in order to know some subjects' characteristics (genre, age, academic degree, employment, residence province, nationality, game play frequency, English level understanding).

Data Analysis Procedure

In order to reduce our items scale into more simple subsets we had proceeded to a factor analysis with pre test results. We made a principal component analysis followed by varimax rotation requesting a 6 component extraction. We opted for varimax rotation requesting a 6-component extraction because it explains a high total variance and it allows a better identification of factors. The six factors obtained explain 66.692% of Total Variance. Our first factor reflects Anger and Disgust basic emotions and explains 22.506% of total variance. This factor includes the following items: Anger, Agitation, Confusion, Contempt, Disgust, Embarrassment, Tension, Guilt, Perturbation, Nervousness, Boringness and Restlessness. The second factor is related to Happiness and explains 10.951% of total variance. It includes Amusement, Contentment, Happiness, Relief, Comfort, and Satisfaction. The third factor reproduces Tranquillity and explains 9.655 % of Total Variance. It comprises Tranquillity, Relaxation and Affection. The fourth factor is connected to Sadness and explains 8.111% of Total Variance. It encloses Pain and Sadness items. The fifth factor enlaces Interest, Calm and Confidence, and explains 7.663% of Total Variance. We had denominated this factor as Non specific emotions because it includes different emotions. The last factor evinces Fear and explains 7.806% of Total Variance. It encircles Fear and Terror.

Our questionnaire internal consistency had been tested by Cronbach's Alpha. Cronbach' Alpha presented a value of 0.874.

By analyzing factors differences between Pre-Questionnaire (pre emotional state) and Post Questionnaire (post game emotional state) results were identified. We had proceeded to Paired-Sample T Test and we had only taken in account differences with statistical significance at a Confidence Interval of 95 %.

[2] All questionnaires are based on Gross questionnaire and STAI.

4 Data Analysis[3]

As we can see, in Appendix 1, there are six factor emotion means and four specific emotion means presented. This is basically due to the Non specific factor, which had presented significant differences by the multiple conditions (games) and consequently we opted for analyzing each emotion separately (Interest, Calm and Confidence) in order to understand which ones have a relevant variation. Surprise appears as part of Sadness factor (on Factor Analysis). However if we attend to emotions theories this could not make sense (e.g. [15]), therefore we had also analyzed Surprise separately.

In a first glance, we can say that we obtain numerous significant differences through the different conditions/videogames, which already suggests that this Storytelling genre elicits emotional variation (see gray areas of Appendix 1). Examination of Appendix 1 reveals that:

The Interest emotion had presented the highest representative reduction on Final Fantasy VII and MGS, which are the less interactive game sequences presented. In contrast, the highest means of that emotion were obtained on Alien Versus Predator 2 and Half-Life 2, which are some of the most interactive game sequences presented.

On Max Payne game we can verify that Anger-Disgust responses are higher than on Pre-test at $p<0.001$, which suggests that it was able to produce Anger-Disgust as pointed by game players inquired. However the mean difference at Anger-Disgust was supplanted by Silent Hill 2 and Doom 3, which we also expected to elicit that kind of emotions. Max Payne was also, besides FFVII and MGS, able to slightly reach the Sadness factor maybe because it makes use of non-interactive sequences.

Games that had obtained the highest significant levels of Sadness were Final Fantasy VII and Metal Gear Solid, respectively, as expected. At the same time, these games presented significant reduction of positive emotions, as Happiness, Tranquillity and Confidence.

Games presented for Happiness purpose (KH and B G&E) did not obtain a relevant increase in relation to the pre-test (at $p<0.05$). However we consider Kingdom Hearts as a game that can be able to elicit the Happiness factor based on four evidences: a) Kingdom Hearts presents an extremely significant reduction of Sadness, the opposite dimensional emotion in the Circumplex model [15]; b) Kingdom Hearts is the only game that does not present a significant reduction of the Happiness factor; c) Kingdom Hearts is the only one that presents an increment of positive valence comparatively to Pre Test, though its difference is not statistically relevant; d) Comparison of Kingdom Hearts to all other games (through factors means differences analysis using Paired-Sample T Test too) reveals that Happiness factor is significantly higher after this game presentation than after all the others.

Games presented for Tranquility and Calm emotions purpose (Ico, Myst III) did not obtain a significant increase of those emotions (at $p<0.05$), although they present the highest means on Tranquility Factor (accompanied by Kingdom Hearts) and they also do not present representative reduction of Calm. When we proceed to comparison of Tranquility mean obtained on Ico and the other games, we find out that a significant statistical difference in relation to all the other games presented except the games that display a positive valence (Kingdom Hearts, Myst III and Beyond Good & Evil).

[3] Our entire statistic data analysis was processed by SPSS program.

Doom 3, Resident Evil, Silent Hill 2 and Alien versus Predator elicited Fear more intensively and with statistical significant difference. All of these games were expected to elicit negative emotions.

Surprise was successfully elicited by Resident Evil, which was selected with that purpose. In contrast, Half-Life 2 elicits an increase of Surprise but the difference is not representative at a Confidence Interval Level of 95%.

According to Valence variation, Kingdom Hearts is the only one that presents an increment of positive valence when compared to the Pre Test. However this difference is not statistically significant.

Finally, we can observe that Surprise, Anger-Disgust and Fear were elicited by some of the videogames presented, which suggest that this spectrum of emotions is able to being elicited by videogames. We can also conclude there is some evidence to affirm that Happiness can apparently be elicited by the game Kingdom Hearts. It took us some trouble to elicit Tranquility, despite the fact that, from all games presented, ICO seems to be the most efficient to do it. Sadness was only elicited through the non-interactive game sequences – Final Fantasy VII, Metal Gear Solid and Max Payne - demonstrating the failure of eliciting this factor with interactive structures.

Other Variables Effects

We have also analyzed the effect of other variables. We used a T Test for Independent samples to analyze Genre, Games Presentation Order and Habit. One-Way ANOVA was used to analyze Game's Play Frequency. We had also assumed a Type I Error of 0.05 on those analyses at the moment of considering results.

An ANOVA test of game play frequency reveals that persons who frequently play games show more Interest and Confidence on Half Life 2. Some possible explanation to that can be found in the increased realistic visual and physical environment comparatively with other games faced by usual players in Half Life 2.

Persons who do not have game playing habits felt higher levels of Surprise in Myst III. The exquisite and at the same time beautiful world depicted in Myst can easily seize new comers to the game world. In part, this can explain the big success of the Myst series among players and above all non-players that usually buy these games and have never finished them [24].

Attending to representative differences in terms of sex, the male subjects always present higher results than female ones on differences verified. Those results represent a difference to a Gross study [6], in which it was concluded "Women reported greater levels of target emotions than men did." This difference can suggest that nowadays games are more able to elicit emotions on men than on women and films to elicit more emotions on women than men.

In respect to order presentation, there are some significant differences, according to which normal order always presents higher results than inverted order. Those results could reflect the "Excitation Transfer" effect refereed by Zillmann [19].

5 Spectrum Comparison Games – Movies

We built circumplex models for our games and Gross films, and distributed them across their emotional quadrants (see Fig.2 and Fig.3). We took into account not only the statistical difference from pre-test to each game but also as Gross did the significant differences between games.

With these emotional models we can see that games are almost as emotionally diverse as films at least in 75% of the emotional spectrum. However according to our research we must argue that games are having problems to hit the other 25%. These 25% are under the basic emotion of Sadness, the *Deactivated-Negative Quadrant*.

If we look at the study performed by Lazzaro [9], even considering that she researched a wider range of game genres, including multi-player gaming, we can also see that the emotions had a similar purpose to the ones evoked by the games - Fear, Surprise, Disgust, Kvell (sort of Pride), Fiero (sort of Fierce), Schadenfreude (sort of Glee) -, it stands only for the two upper quadrants. Plus, she also found that the "relaxing effects of games are very appealing and some apply its therapeutic benefits to 'get perspective', calm down after a hard day", arguing here for the existence of the eliciting of calm and relaxing emotions and so the activation of the de-activated positive quadrant. She ended reaching the same emotional spectrum as we have, albeit with considerable different weights in emotional dimensions.

The lacking of one entire quadrant within the Interactive Storytelling media raises great problems because it makes the creation of a wide range of themes and contents impossible through it. Most of our understanding of the Human Condition is *narrativised* making use of emotions of this quadrant once they are responsible for most of the inner problems we face in daily life. "Sadness focuses the person on him- or her- self [..] the self focuses on the consequences of not achieving its goals"[2].

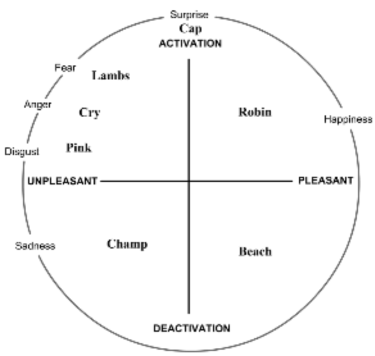


Fig. 2. Movie Circumplex Model

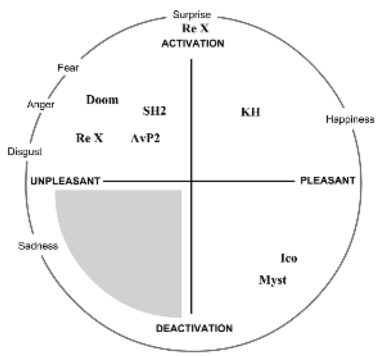


Fig. 3. Game Circumplex Model

6 Conclusions and Perspectives for Future Work

We can conclude that our results strongly suggest that Virtual Storytelling has emotion elicitation capacities and it is possible to produce interactive stories with several emotional abilities.

Our results suggest that interactive stories are capable of eliciting Anger-Disgust, Fear and Surprise efficiently. We faced difficulties in showing the same interactive stories' capacities concerning to Happiness and Tranquility. Nevertheless, we consider that there is some evidence that Happiness is possible to be elicited by games. We believe that these difficulties can be related to our concern to create a pleasant and comfortable initial condition so that the external subject's emotional conditions do not interfere with our experiment results. According to that our Happiness and Tranquillity initial scores are very high and were difficult to be supplanted by any game. This concern could be faced as a handicap of our study. Despite this, other studies had pointed out that both Happiness and Tranquillity in the Relax emotion form could be elicited by games (e.g. [9]) as we have already mentioned.

As we did not find any interactive game sequence able to elicit Sadness, despite our efforts, we consider that it can suggest that games are unable to elicit the Sadness Factor emotions. We noticed that when games used the same non-interactive mechanics film language did, they were able to elicit that emotion (as we can see on FFVII, Max Payne and MGS Sadness results).

The two most important conclusions that justify the necessity of the diversified emotional spectrum come from the analysis of the Interest Emotion and the analysis of the Sex variable. Based on our Interest level results we realized that subjects do not apparently like non-interactive games as they like the interactive ones. FFVII and MGS were almost no interactivity sequences, making use of film language asking for the passivity of viewing instead of acting upon them, and so they showed the lowest interest levels from the 14 sequences. From the analyses of the sex variable, we find results demonstrating overall male higher results than female ones in opposition to Gross's [6] results with movie clips eliciting. This suggests that games could be more capable of eliciting emotions on men than on women. This is not a big surprise to our study as, since the moment we submitted our subjects to the fact that they should have interest in games to enter the test, we ended with 81.8% men and only 18.2% women doing our tests. Surprising is the fact that these 18.2 % were female with high interest in doing the games tests, and even their levels hit lower than men's.

We have to conclude from this, that the storytelling strategies used were surely different from those used in the movie clips of Gross tests, and that they were almost certainly produced having in mind a men target.

This means that we'll need to invest more specifically in the building of new narrative methodologies for the creation of interactive stories with the ability to produce Sadness emotion. In our view and according to our data results, the incapacity of producing Sadness interactively raised two big problems to the interactive stories till now: a) low demographics, above all from the female population; b) diminishing of interest and so a consequent lower engagement of the overall population through the introduction of non-interactive moments.

Acknowledgments

This research is performed in the frame of the INSCAPE Integrated Project (EU RTD contract IST-2004-004150), which is funded by the European Commission under the sixth Framework Programme. More information on the project is available at www.inscapers.com

References

1. Barret, L.F. & Russell, J.A. (1999), The structure of Current Affect: Controversies and Emerging Consensus, Current Directions in Psychological Science, Vol. 8, Nº 1.
2. Barr-Zisowitz,C. (2000), "Sadness" – Is there such a Thing?, In Lewis, M. & Haviland-Jones, M. (Eds.), Handbook of Emotion, New York: Guilford Press.
3. Damásio, A., (2000), O Sentimento de Si: O Corpo, a Emoção e a Neurobiologia da Consciência, Mem Martins: Publicações Europa-América.
4. Ekman, Paul, (1993), Facial Expression and Emotion, American Psychologist, Vol. 48, No. 4, 384-392.
5. Frijda, N. H. (2000), The Psychologists' Point of View, In Lewis, M. & Haviland-Jones, M. (Eds.), Handbook of Emotion, New York: Guilford Press.
6. Gross, J. J. & Levenson, R. W, (1995), Emotion Elicitation Using Films, Cognition and Emotion, 1995, 9 (1), 87 –108, http://www-psych.stanford.edu/~psyphy/publications.htm
7. Izard, C. E. (1991), The Psychology of Emotions, New York: Plenum.
8. Lazarus, R. S. & Smith, C. A. (1988), Knowledge and Appraisal in the cognition-emotion relationship, Cognition and Emotion, 2.
9. Lazzaro, Nicole, (2004), Why We Play Games: The Four Keys to Player Experience, XEODesign, presented at the Game Developers Conference 2004
10. Microsoft (2005), Robbie Bach, J Allard, Peter Moore: Electronic Entertainment Expo (E3) 2005, transcripts, http://www.microsoft.com/presspass/exec/rbach/05-16-05E3.mspx
11. Oatley, K. & Johnson-Laird, P. N. (1987), Towards a cognitive theory of emotions, Cognition and Emotion, 1.
12. Panksepp, J. (2000), Emotions as Natural Kinds within the Mammalian Brain, In Lewis, M. & Haviland-Jones, M. (Eds.), Handbook of Emotion, New York: Guilford Press.
13. Plutchick, R. & Kellerman (1989), Emotion: Theory, Research and Experience, vol. 4, The measurement of Emotion, San Diego: Academic Press.
14. Rottenberg, J., Ray, R. & Gross, J. (in press), Emotion Elicitation Using Films, In Coan, J. & Allen, J. (Eds), The handbook of Emotion elicitation and assessment, New York
15. Russell, J., Lemay, G., (2000), Emotion Concepts, In Lewis, M. & Haviland-Jones, M. (Eds.), Handbook of Emotion, New York: Guilford Press.
16. Sherer, K.R. (1984), Emotion as a multicomponent process: A model and some cross-cultural data, Review of personality and social psychology.
17. Vaz Serra, A. (1999), O Stress na Vida de Todos os Dias, Coimbra.
18. Zagalo, N., Branco, V., "Storytelling and Interactivity in Videogames, from MYST to ICO", in Deliverable 3.1.1 of INSCAPE Integrated Project (EU RTD IST-2004-004150)
19. Zillmann, D., (1991), Empathy: Affect from bearing witness to the emotions of others. In: J. Bryant and D. Zillmann (eds.), Responding to the screen: Reception and reaction processes, 135-167. Hillsdale, N J: Erlbaum.

Appendix 1. Means Emotion Intensity of each game for each emotion and Paired Samples T Test results (of Pre Test comparisons)

Emotions \ Conditions	Pre	Max	FFVII	Ico	AvP 2	SH2	HL2	Myst	Doom	KH	Medal	ReX	MGS	BGE	Postal
Anger-Disgust Factor	**14,24**	**29,36** t=-5,70, p=,000	**20,91** t=-2,61, p=,014	**10,42** t=1,83, p=,077	**26,48** t=4,85, p=,000	**30,48** t=-6,06, p=,000	**18,27** t=-1,39, p=,174	**12,15** t=,919, p=,365	**33,00** t=-6,83, p=,000	**11,06** t=1,4 , p=,171	**26,55** t=-5,17, p=,000	**25,79** t=-4,8 , p=,000	**15,45** t=-,497 , p=,623	**9,52** t=2,01, p=,053	**20,42** t=-2,11, p=,043
Happiness Factor	**23,15**	**15,76** t=3,78 , p=,001	**12,03** t=6,1, p=,000	**17,64** t=2,86, p=,007	**14,06** t=4,67, p=,000	**7,33** t=9,56, p=,000	**18,18** t=2,61, p=,014	**14,12** t=5,55, p=,000	**12,88** t=4,84 , p=,000	**22,79** t=,189 , p=,851	**12,42** t=5,18, p=,000	**6,76** t=10,55 , p=,000	**6,52** t=11,61 , p=,000	**18,58** t=2,54, p=,016	**16,52** t=2,72 , p=,010
Tranquillity Factor	**15,09**	**6,33** t=9,31 , p=,000	**10,18** t=5,19, p=,000	**12,27** t=2,84, p=,008	**4,76** t=12,3, p=,000	**3,91** t=11,0, p=,000	**7,36** t=9,27, p=,000	**11,48** t=4,21, p=,000	**3,88** t=10,99 , p=,000	**12,52** t=2,55 , p=,016	**4,73** t=10,21 , p=,000	**2,97** t=13,47 , p=,000	**8,82** t=5,91 , p=,000	**11,12** t=4,55, p=,000	**6,73** t=8,17 , p=,000
Sadness Factor	**1,82**	**3,55** t=-2,82 , p=,008	**4,24** t=-3,6, p=,001	**0,52** t=3,11, p=,004	**1,06** t=1,72, p=,096	**1,42** t=,76, p=,451	**0,94** t=1,95, p=,059	**0,55** t=3,08, p=,004	**1,33** t=1,05 , p=,301	**0,76** t=2,24 , p=,033	**1,85** t=-,05, p=,957	**2,18** t=-,63 , p=,536	**3,73** t=-3,63 , p=,001	**0,64** t=2,44, p=,020	**1,61** t=,33 , p=,743
Non specific emotions	**16,36**	**10,42** t=6,89 , p=,000	**9,85** t=7,15, p=,000	**12,64** t=4,21, p=,000	**9,73** t=8,04, p=,000	**6,94** t=11,1, p=,000	**11,42** t=6,32, p=,000	**12,12** t=4,44, p=,000	**8,30** t=8,3 , p=,000	**12,24** t=4,76 , p=,000	**8,67** t=8,41, p=,000	**6,48** t=11,35 , p=,000	**8,61** t=8,35 , p=,000	**11,39** t=6,61, p=,000	**8,30** t=8,02 , p=,000
Fear Factor	**1,24**	**7,24** t=-5,7 , p=,000	**2,27** t=-1,48, p=,149	**0,82** t=,815, p=,421	**8,00** t=-5,68, p=,000	**8,36** t=-6,03, p=,000	**4,18** t=-3,06, p=,004	**0,94** t=,585, p=,562	**10,27** t=-7,23 , p=,000	**1,24** t=0 , p=1	**6,03** t=-5,71, p=,000	**9,36** t=-6,85 , p=,000	**1,58** t=-,562 , p=,578	**1,00** t=,421, p=,677	**3,33** t=-2,58, =,015
Interest	**5,91**	**4,61** t=3,15 , p=,004	**2,61** t=9,17, p=,000	**3,91** t=5,79, p=,000	**4,67** t=3,29, p=,002	**3,30** t=6,46, p=,000	**4,82** t=2,95 , p=,006	**4,00** t=5,69, p=,000	**4,09** t=3,76 , p=,001	**4,39** t=4,31 , p=,000	**4,52** t=4,4 , p=,000	**3,24** t=5,88 , p=,000	**2,70** t=9,1 , p=,000	**4,00** t=6,31, p=,000	**3,58** t=5,03 , p=,000
Surprise	**2,94**	**3,39** t=-,961, p=,344	**2,52** t=,93, p=,358	**2,79** t=,333, p=,742	**3,61** t=1,46, p=,153	**3,42** t=-,889, p=,381	**3,55** t=-1,07, p=,295	**1,94** t=2,33, p=,026	**3,97** t=-1,98 , p=,056	**3,00** t=-,135, p=,893	**3,03** t=-,215, p=,831	**5,09** t=-3,72 , p=,001	**1,52** t=3,39 , p=,002	**2,12** t=2,26, p=,031	**3,36** t=-,951, p=,349
Calm	**5,15**	**4,61** t=6,62 , p=,000	**4,24** t=1,9, p=,067	**4,64** t=1,41, p=,167	**1,91** t=8,43, p=,000	**1,70** t=8,66, p=,000	**3,00** t=6,13 , p=,000	**4,67** t=1,18, p=,247	**1,52** t=9,73 , p=,000	**3,76** t=3,58 , p=,001	**1,58** t=9,36 , p=,000	**1,61** t=9,87 , p=,000	**3,55** t=3,60 , p=,001	**4,00** t=3,16, p=,003	**2,21** t=8,65, p=,000
Confi-dence	**5,30**	**3,27** t=4,96 , p=,000	**3,00** t=5,97, p=,000	**4,09** t=2,82, p=,008	**3,15** t=5,36, p=,000	**1,94** t=8,07, p=,000	**3,61** t=3,71, p=,001	**3,45** t=4,43 , p=,000	**2,70** t=5,82 , p=,000	**4,09** t=2,94 , p=,006	**2,58** t=6,32 , p=,000	**1,64** t=10,91 , p=,000	**2,36** t=6,59 , p=,000	**3,39** t=5,31, p=,000	**2,52** t=6,65 , p=,000

Every salient means (gray areas) represents differences with statistical significance (at a Confidence Interval of 95 %) related to Pre Test Means.

Acronyms used throughout the text: (*KH*)=Kingdom Hearts; (*BGE*)=Beyond Good and Evil; (*Max*)=Max Payne; (*FF VII*)=Final Fantasy VII; (*Medal*)=Medal of Honour Allied Assault; (*HL2*)=Half-Life 2; (*SH2*)=Silent Hill 2; (*Re X*)=Resident Evil: Code Veronica X; (*AvP 2*)=Alien Versus Predator 2; (*MGS*)=Metal Gear Solid; (*Robin*) = Robin Williams in Concert; (*Beach*) = Beach scene; (*Cap*) = Capricorn One; (*Lambs*) = Silence of the Lambs; (*Cry*) = Cry Freedom; (*Pink*) = Pink Flamingo; (*Champ*) = The Champ; *(Doom)* = Doom 3; *(Myst)* = Myst 3

The Control of Agents' Expressivity in Interactive Drama

Nicolas Szilas[1,2] and Maurizio Mancini[1]

[1] LINC, IUT de Montreuil, 140 rue de la Nouvelle France,
93100 Montreuil, France
m.mancini@iut.univ-paris8.fr
[2] Department of Computing, Macquarie University,
NSW 2109, Australia
nicolas@ics.mq.edu.au

Abstract. This paper describes how conversational expressive agents can be used in the context of Interactive Drama. This integration requires some automatic tagging of the generated text, according to the current dramatic situation. Experimental results of the implemented prototype are presented.

1 Agents and Interactive Drama

1.1 Project Description

Interactive Drama (ID) is a new narrative form which takes place on a digital medium and where the audience plays an active role by deciding how to behave in the story, acting as a character. People are accustomed to "fake" ID with Interactive Fictions or adventure video games. In those cases, the freedom of the user as a character is very limited compared to the range of actions that characters are supposed to have in a story. True ID has not yet been reached but several prototypes are going in this direction [4], [7], [12], [14], [18].

Part of the difficulties of ID lies in the fact that it requires many diverse components. Furthermore, the research in ID is scattered and systems have been developed rather independently so far, with limited reuse of components developed externally. This paper aims at combining two components to create a more compelling ID: an expressive Embodied Conversational Agent (ECA) system called Greta [9] and an interactive narrative engine called IDtension [14]. The primary goal of this integration is to convey the story with emotion through using emotionally expressive agents.

Even if most narrative models focus on the cognitive dimension of storytelling, emotion is now recognized as a key feature of narrative and drama. Regarding the narrative film in particular, the narrative structure itself is constructed in order to produce some emotions such as hope or fear [16]. According to Noel Carroll, emotions even constitute the condition of film understanding. They allow the viewer to focus his/her attention on important elements in the narrative [3].

We assume that these general data about emotion in narrative still hold in the interactive case. ID should trigger the user's emotions and this paper investigate one of the main sources of emotion in Drama: the emotions expressed by the characters.

In fact, emotional agents have been the basis of ID since the birth of the field [6], but our project differs from previous systems for the following reasons:

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 115 – 124, 2005.

- The story is dynamically generated by a narrative engine rather than by a simulation of agents [15]. Thus, we raise the theoretical and practical issue of how to coordinate emotions at both narrative and behavioral levels;
- The story events themselves are generated (including the texts that are used in dialogs), which prevent any scripted solution for emotion generation;
- The expressive agent component enables the system to generate precise and rather realistic facial expressions, which opens the way to more engaging ID.

1.2 Generative Interactive Drama: IDtension

IDtension is a long term research project, started in 1999, which aims at solving the paradox of interactive narrative (see [14], [15] for details). It is built around a narrative engine which dynamically generates a story and lets the user choose the actions of one character. It operates at the logical level, in a medium independent manner. It contains the following four main distinctive features:

1. The story model is a fine grain model, in the sense that it manipulates elementary actions rather than larger units like for instance beats [7] or scenes. Such a fine grain model provides more interactivity but prevents the author to easily craft each scene. These scenes are generated by the system.

2. The system includes a model of the user aimed at estimating the impact of each possible action on the user according to several narrative criteria [14]. Some of these criteria focus on believability. Other criteria such as *complexity* and *conflict* are only guided by narrative concerns. Some of them are related to the user's emotions: the system is able to trigger one action instead of another with the intention to trigger a given emotion in the user. Conflict for example occurs when the user's character has a specific goal while the mean to reach that goal (a *task*) goes against his own ethical values. Such conflict is the core of classical drama and produce strong emotional reaction in the viewer.

3. The articulation of actions is twofold. IDtension considers generic actions and specific tasks. Generic actions stem from narratology [2], [17]. They are, for instance, inform, encourage/dissuade, accept/refuse, perform, felicitate/condemn. Tasks are specific to a story: kiss, hug, slap (in a romance story) or threaten, torture, kill (in a *roman noir)*, etc. This makes it possible to handle complex actions like “John tells Mary that Bill has robbed her jewels” without requiring from the author to explicitly enter them into the system. The narrative engine handles logical forms such as *Inform (John, Mary, have_finished(Bill, rob , [jewels,Mary])*. The author only specifies the task (rob), the characters (Mary, John, Bill) and the objects (the jewels) in the story.

4. IDtension explicitly processes the notion of (ethical) values. Values are thematic axes according to which each task is evaluated. Such values include honesty, friendship, family, etc. This mechanism adds beyond the pure performative dimension, another dimension to the story namely the axiological dimension. The *Model of the User* processes those values to evaluate some narrative criteria, in particular *conflict*.

Action selection is performed in three steps:

1. The *Narrative Logic* generates the set of all possible actions at a given time in the narrative via a set of narrative rules.

2. The *Model of the User* assesses each of these actions according to its estimated impact on the user. The actions can then be ranked.
3. The *Theatre* displays the selected action by generating the text with a template based approach and by sending it to the virtual characters. The system alternates actions chosen by the engine and actions chosen by the user.

1.3 Expressive Embodied Conversational Agents: GRETA

ECAs are entities with a human-like appearance capable of taking part in dialogs with the users. They communicate multimodally with their human interlocutors (or with other agents) through voice, facial expression, gaze, gesture, and body movement.

Fig. 1. The Greta agent

Greta (see Fig. 1) is an ECA system that given an XML-tagged text as input provides some animation files that drive an embodied representation of a virtual human as described in [9]. The input file contains both the text that the agent has to speak and the information on the communicative intent of the conversation, defined using an XML-based language called APML (Affective Presentation Markup Language). APML is based on the taxonomy of communicative functions defined in [11] by Poggi. The tags of APML provide information on the Speaker's Mind -on his/her beliefs, goals and emotions:

– belief tags — they convey information about the certainty/uncertainty of the beliefs that the Speaker is referring to and their source (for example, long-term memory).
– goal tags — they refer to the Speaker's goals, for example the generic goals of giving general information. Or they can help the Speaker underline the most important part of the sentence, specify logical relation between different parts of the discourse, manage the evolution of the conversation (for example turn-taking).
– emotion tags — they are meant to convey the affective states of the Speaker that will mainly influence the facial expression and body gestures that the agent will choose to show.

In the example displayed in Table 1, we can distinguish the *performative* tags on lines 2 and 17 that represent the goals the agent wants to achieve with the two sentences inside of them. On line 9 and 19 the presence of the *rheme* tags means that the surrounded text is the important part of the discourse. Finally, some emotional content ("resentment") is specified on line 19.

Moreover, APML provides some special tags and attributes to drive the speech synthesizing process. For this purpose, the tag *boundary* (lines 14, 25) and the attribute *x-pitchaccent* (lines 4, 11, 21) have been introduced, as explained in [13]:

- boundaries : they are perceived as prosodic phrase-final rising pitch and/or lengthening.
- x-pitchaccents : they are perceived as word-level stress, accent or emphasis.

Table 1. A typical APML input file for the Greta agent

```
1. <apml>
2. <performative type="warn">
3. <theme>
4. <emphasis x-pitchaccent="Hstar">
5. Please
6. </emphasis>
7. do not tell me what works
8. </theme>
9. <rheme>
10.for
11.<emphasis x-pitchaccent="Hstar"
   deictic="self">
12.me
13.</emphasis>
14.<boundary type="LL"/>
15.</rheme>
16.</performative>
17.<performative type="suggest">
18.Would you just please
19.<rheme affect="resentment">
20.mind your
21.<emphasis x-
   pitchaccent="Hstar"
   adjectival="foreign">
22.own
23.</emphasis>
24.business
25.<boundary type="H"/>
26.</rheme>
27.</performative>
28.</apml>
```

Given an APML file, the Greta system looks in a dictionary of signals for the gestures/facial expressions associated to the communicative functions specified by the tags in the input file. A gesture/face planner is responsible for instantiating these signals according to temporal and contextual constraints (e.g. coarticulation between gestures/facial expressions). *Festival* (Univ. of Edinburgh) is used for speech synthesis.

The exact procedure the agent performs in generating animation is:

1. Input — the starting point is the APML file with tags representing the communicative functions listed above (beliefs, goals, emotions).
2. Signals selection — the system looks for the associated gesture/facial expressions to each tag in the signal dictionary.
3. Synchronization — the signals are instantiated according to temporal and contextual constraints (e.g. coarticulation). Each tag of the APML file corresponds to some text (the text which is "nested" in the tag). So from the duration of the speech associated to the text tags, the exact starting and ending times of each tag (and consequently for each signal) are calculated. At this point the correct temporal characteristics of the facial expressions [1] and gestures [5] are generated.

2 Theoretical Approach

2.1 A Communication-Based Model for Expressivity

The relationship between agent's expressivity and narrative expressivity is as complex as the relationship between character and narrative.

One should be careful not to reduce a narrative to the simulation of a fictional world. A narrative contains only a limited number of actions usually discontinuated simply evoking a fictional world in the receiver but does not contain a simulated world. In addition these limited actions are carefully crafted to convey a certain mean ing, to trigger certain emotions, which goes beyond a world simulation. Hence the famous Hitchock's quote: “Drama is life with the dull bits cut out”.

Research in embodied agents is usually performed (and funded) in a context of a world simulation, in order to simulate humans (“virtual humans”, “lifelike characters”, etc.). Research in emotional agents in particular aims at providing mechanisms for an agent to react properly to given situations by expressive realistic emotions. "Believable agents" [6] aim at refocusing the simulation towards its expressivity. However, the actions of believable agents are still resulting from the simulation of humans in the environment even if those mechanisms tend to be more expressive (e.g. exagerated).

According to the general view of drama introduced above (drama is more and less than a world simulation), we meet two sets of problems in directly applying virtual human research to ID:

– Simulating a set of emotional believable agents and then selecting only the parts that are shown to the user seems to constitute quite a detour. This simulation is a challenge in itself.
– Integrating expressive features calculated by the narrative engine with those calculated according to the world simulation is problematic.

For those reasons we decided to focus only on what is perceived by the user, namely the actions on the screen. The actions calculated by the narrative engine come with the necessary information to calculate the corresponding expressive features. In particular, those actions are calculated according to a goal structure, which constitutes a key feature in emotions [8].

We call this approach a communication-based model, because it links emotions to actions that are conveyed to the user rather than calculating it according to the agent's states. Between two actions, no emotion is calculated. This previous affirmation is hard to follow in a virtual environment because characters are visible between narrative actions. The solution is to assign to an agent the emotion of the previous narrative event it was involved in with a decreased intensity in time.

2.2 Authoring Expressivity

In classical (non interactive) drama, the expressivity of a given sentence is produced by the actor based on the annotated script and with the help of the director. In the digital context such human skill is not available during the performance. To compensate two actions are made available:

– manually annotating each sentence in detail, so that the agent "only" follows gestural and prosodic directions;
– calculating on the fly these expressive directions according to a set of predefined linguistic rules and data.

In the context of generative Interactive Drama, a concept such as "each sentence" is nonexistent. These sentences are generated on the fly, and even if their number is finite, it would be much too large to allow any manual annotation. Only templates are available (sentences with gaps to fill), which makes expressivity authoring uneasy, because each piece of text could appear in different contexts. Moreover, this kind of annotation using the APML language requires a high level of expertise in linguistics as well as specific training [13].

However, relying solely on automatic tagging of text does not provide the best quality of expressivity. Doing so would furthermore frustrate the author who would have no control over agent's expressivity.

Thus, a hybrid approach is suggested. The system automatically creates expressive annotations of the text, while the author is able modify to a certain extent the expressive rendering. The author's work would require some easy tagging of the text, which will serve as a basis for APML generation. This strategy has been partially implemented, as described in Section 3.4.

3 Implementation

3.1 Technical Integration

Some modifications of the Greta system have been made, in order to integrate it with the IDtension system. A TCP/IP server able to accept external connections has been added. It has been implemented as a parallel thread that constantly waits for incoming TCP/IP connections from any external source (client). In this manner another process (that is, the IDtension system) can communicate to Greta and send data as text strings.

The IDtension system decides which APML data has to be transmitted by taking the logical form of an action and generating the corresponding text to be spoken with the tags, according to the rules described below. It then connects to the Greta's server and sends the APML data through the TCP socket as a sequence of strings. Each time the sending of APML data is completed, the Greta system elaborates it as described in section 1.3 and generates the corresponding graphical animation.

3.2 Rules for the Tagging of Emotions

Two types of emotions are handled in the narrative engine: emotions that are associated to the user and managed in the *Model of the User* and emotions that are associated to the characters and managed by the *Theatre*. Only this second type of emotions will be considered. In addition, only emotions associated with the character's current action are considered, not those triggered in reaction to other character's actions.

The method of emotion association is related to the cognitive structure of emotions described in [8], but instead of identifying emotional states of the agents, the system focuses on actions, as suggested in [10]. Seven emotions have been implemented so far (see Table 2).

Each automatic emotional tagging is defined by the following elements:

- types of actions the emotion can be associated to,
- condition of triggering the emotion,
- calculus of the intensity.

Two of the these emotions will be detailed here: worry and disappointment.

Worry: this emotion occurs when the character perceives that there is a chance of failure of the agent's performance.

The actions for which worry could be expressed are all actions related to a task before it is accomplished (see Table 1).

The “worry” emotion is triggered if all the following conditions are met:

- The importance of the goal for the character is superior to a given threshold.
- The perceived risk cumulated over all known obstacles for the corresponding task is superior to a given threshold.

  Note that these data are available in the narrative model [14].

The intensity is calculated differently according to the action:

– if the action is an "inform hinder", which means that the character is talking about a specific obstacle on the task:

$$\text{intensity} = \text{importance(goal)} \times (1\text{-perceived\_risk(obstacle)})$$

– in all other cases:

$$\text{intensity} = \text{importance(goal)} \times \Pi_i(1\text{-perceived\_risk(obstacle}_i))$$

where i ranges in all known obstacles.

Disappointment: this emotion occurs when an obstacle has been met.
It is expressed with the action "inform have_been_blocked", when a character informs another that s/he met an obstacle during the execution of a task.
The "disappointment" emotion is triggered if all the following conditions are met:

– The importance of the related goal for the agent is superior to a given threshold.
– The obstacle was met recently, that is the difference between the current date and the date of the triggering of the obstacle is below a given threshold.

The intensity equals the importance of the goal for the character.

Table 2. Emotions implemented in the system

Type of emotion	Emotion	Types of action involved
before task performance	enthusiasm	inform {want, have_begun}
	worry	inform {want, have_begun, hinder}; accept
after task performance	satisfaction	inform {have_finished}
	disappointment	inform {have_been_blocked}
after task performance, value-related	disgust	inform {can,want, have_begun}
	anger	dissuade, condemn
	shame	inform {have_begun, have_finished}

3.3 Rules for the Tagging of Boundaries

Adding the boundary tags requires three steps:

– segment the text,
– choose between theme and rheme,
– choose between the different types of boundaries.

These steps when performed by a human are quite difficult. They require some knowledge about theories on English prosody as well as practice [13]. Therefore, the rules proposed below are a strong approximation of the proper coding. We intend to improve these rules in the future.

To segment the text, two rules are used:

- Each new predicate corresponds to a new segment. For example, the action *inform (Joe, Mary, want(Joe, swim))*. ("Joe tells Mary that he wants to swim") is broken up into two segments: "Joe tells Mary that" and "he wants to swim".
- Each punctuation marks a new segment.

The choice between *theme* and *rheme* is made based on the two following rules:

- The larger segment of a sentence which correspond to a single action is a rheme (excluding information).
- For information, the content of the information is a rheme (*what* is informed) and the rest is a theme. This corresponds to the fact that what is new in a given information is the content, not the surrounding text. For example, in the sentence: "You know what? Joe bought a car", the segment "You know What?" is the theme, while the segment "Joe bought a car" is the rheme (new information).

The choice of the type of boundaries has been limited to two types of boundaries: LH (rising pitch) and LL (falling pitch).

The following rules are applied [13]:

- For boundaries related to a predicate, if it is a theme the type "LH" is given, and if it is a rheme, the type "LL" is given.
- For boundaries related to the punctuation, the type is "LL" if it is a "period" and "LH" for all other punctuation marks.

3.4 Rules for the Tagging of Emphases

The tagging of emphases is performed in two steps:

- detection of logical components of the action that should be emphasized,
- choice of a method for emphasizing.

The components to which an emphasis can possibly be applied are characters, objects, places, goals, tasks and obstacles. The decision whether an emphasis should be applied or not is related to the need to express some contrast on this component [13]. A component is contrastive when it brings some new information to the addressee. For example, consider the following dialog:

Joe to Mary: "I want to <u>swim</u>"
Mary to Joe: "If you swim, be careful then, <u>it's dangerous</u>"

In the first line, the new information is "swim", but in the second line, the new information is "it's dangerous".

The narrative engine manages contextual information in terms of narrative sequences or processes [14]: to most actions is associated an "initiating action" to which the current action is responding (which is not necessary the action played previously).

Once the decision to add an emphasis to a component has been made, the next step is to surround the corresponding text with the emphasis tags. However, for complex components putting stress on a long piece of text is not adviced. For example, in the sentence "to evacuate the building, you should trigger the fire alarm" the emphasis should be put on "fire alarm" not on "trigger the fire alarm". This choice is difficult to make automatically because it requires either some knowledge about the real world ("fire" and "alarm" are dramatic elements) or linguistic knowledge ("trigger" is a verb and "the" an article, while "fire alarm" is a substantive).

We have chosen to solve this issue by letting the author define which words in the text s/he writes should be emphasized. In the template for the task *triggerFireAlarm*, the author puts "trigger the *fire alarm*". Finally, the rule is as follow: if the component contains an author-defined emphasis, then only the parts defined by the author should be emphasized.

4 Results

An interactive scenario named "Mutiny" has been implemented, which is about four prisoners trying to escape from an old galleon by starting a riot. The user plays "Jack", one of the prisoners. S/he interacts by selecting an action into a large list. Actions are displayed with both text and Greta's head (body movement is also automatically generated by Greta, but is not displayed). Fig. 2 represents a possible path that displays four different emotions. Boundaries and emphases are not represented but can be heard in the synthesized speech and seen in the animation's dynamics.

you try to open the gate between the jail and the deck. you don't go very far with this closed gate....
Jack to Lord: "I could not start a riot against Commandant Nelson: it was impossible to open the closed gate."
Lord to Jack: "If I were you, I would definitely start a riot against Nelson..."
Lord to Jack: "My dear Jack, If you start to start a riot against Nelson, then be careful: the guards are armed and ready to stop anyone trying to escape...You need allies... "
you decide to have the hook.
you ask Rak: "have you got the hook ?"
Rak to Jack: "Hey man... I have the hook"
you quietly rob the hook to Rak.
Jack to Clara: "I did it. I managed to rob the hook to Rak ".
Clara to Jack: "What! You dared to rob the hook to Rak ?! What a shame!!".

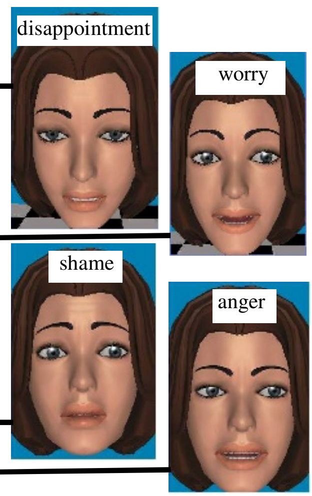


Fig. 2. Text and screenshots of one specific interactive session, highlighting four facial expressions. Note that only one face model is used, even if different characters are involved (Jack for disappointment and shame, Lord for worry and Clara for anger). The emotions of shame and anger are related to characters' ethical values (see Section 1.2).

5 Conclusion

A system which integrates a highly generative ID engine (IDtension) with an ECA (Greta) has been described. This opens the way to more engaging ID where characters' facial and body animations are not pre-scripted but generated based on the current dramatic situation. To reach that goal, we have adopted a communication-based approach in order to focus on action, the core of drama.

As an extension of this research, we intend to study in detail the relationship between the emotions displayed by the agents and to the emotions triggered in the user.

Thus, the agent's expressivity will be generated not only by their in-fiction behaviour but also by the narrative effects calculated by the narrative engine. Such a study would benefit from the data available from Film Studies on how the audience feels empathy or counter-empathy towards the characters in a movie.

Acknowledgments

This research is supported by the *Maison des Sciences de l'Homme de Paris Nord* and by the *Australian Research Council* (Linkage Grant #LX560117).

References

1. Bevacqua, E., Mancini, M., Pelachaud, C.: Speaking with Emotions AISB 2004 Convention: Motion, Emotion and Cognition, University of Leeds, UK, (2004)
2. Bremond, C.: Logique du récit. Seuil, Paris (1974)
3. Carroll, N.: Beyond Aesthetics. Cambridge University Press, Cambridge (2001)
4. Crawford, C: Assumptions underlying the Erasmatron interactive storytelling engine. In Papers from the AAAI Fall Symposium on Narrative Intelligence, Technical Report FS-99-01. AAAI Press, Menlo Park (1999)
5. Hartmann, B., Mancini, M., Pelachaud, C.: Formational parameters and adap-tive prototype instantiation for mpeg-4 compliant gesture synthesis. In Proceed-ings of the Computer Animation 2002. IEEE Computer Society (2002) 111
6. Kelso, M., T., Weyhrauch, P., Bates, J.: Dramatic Presence. TR CMU-CS-92-195, Carnegie Mellon University, Pittsburgh (1992)
7. Mateas, M., and Stern, A.: Towards Integrating Plots and Characters for Inter-active Drama. In Dautenhahn K. et al. (eds): Socially Intelligent Agents. Kluwer Academic Publishers, Norwell, Dordrecht (2002) 221-228
8. Ortony A., Clore G., L., Collins A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge (1988)
9. Pelachaud, C., Bilvi, M.: Computational model of believable conversational agents. In Huget, M.P., ed.: Communication in Multiagent Systems. Volume 2650 of Lecture Notes in Computer Science. Springer-Verlag (2003) 300–317
10. Poggi, I.: Emotions.Un modèle en termes de buts et de croyances. Unpublished seminar, Montreuil, France, July, 12, 2004 (2004)
11. Poggi, I.: Mind markers. In M. Rector, I. Poggi, N.T., ed.: Gestures. Meaning and use. University Fernando Pessoa Press, Oporto, Portugal (2003)
12. Sgouros, N. M.: Dynamic Generation, Management and Resolution of Interac-tive Plots. Artificial Intelligence, 107,1 (1999) 29-62
13. Steedman, M.: Using APML to Specify Intonation. Unpublished Tutorial Paper (2004). ftp://ftp.cogsci.ed.ac.uk/pub/steedman/apml/howto.pdf
14. Szilas, N.: IDtension: a narrative engine for Interactive Drama. In Göbel et al. (ed): Proc. TIDSE'03. Frauenhofer IRB Verlag (2003)
15. Szilas, N.: Interactive Drama on Computer: Beyond Linear Narrative. In Papers from the AAAI Fall Symposium on Narrative Intelligence, Technical Report FS-99-01. AAAI Press, Menlo Park (1999) 150-156
16. Tan, E.: Emotion and the structure of narrative film. Film as an emotion machine. Erlbaum, Mahwah (NJ) (1996)
17. Todorov, T. Les transformations narratives. Poétiques,3 (1970) 322-333
18. Young, R. M., Riedl, M. O., Branly, M., Jhala, A., Martin, R.J. Saretto, C. J.: An architecture for integrating plan-based behavior generation with interactive game environments, Journal of Game Development, 1, 1 (2004)

Agency and the "Emotion Machine"

Josephine Anstey

University at Buffalo

Abstract. In interactive fiction or drama the author's role in manipulating the user into dramatic situations is as important in the construction of the user's sense of "agency," as concerns about the her freedom and choice.

1 Agency, Authorship, Manipulation

When considering interactive fiction, agency has been typically understood as the user's sense that she can act in a virtual environment and that her actions are effective [6]. Choice and freedom for the user are at a premium; narrative engines must be devised that can devise plots on the fly. However, some researchers reject both the possibility and need for absolute freedom in an interactive narrative. [5][4]. This school of thought focuses on producing a more strictly authored narrative which uses physical or social context to constrain the users' actions and controls pacing and surprise. Fiction and drama move their audiences by manipulating their knowledge and emotions [7]. My own work in interactive fiction suggests that rather than looking for ways to provide freedom and choice, we should be researching dramatic methods for manipulating users that work successfully in an interactive context.

Semioticians suggest that a reader/spectator is an agent because it is the work she does that creates meaning[3]. This suggests that for authorial manipulations to work, the user has to go along with them. She has to willingly suspend disbelief and actively engage her cognitive and emotional self in the construction of the story. On the other side, the author has to be able to reliably evoke narrative reasoning and emotion. Bernard Perron applies Ed Tan's expression "emotion machines", originally used to describe movies, to video games and I suggest that interactive drama should also be an emotion machine. I believe the role of the author is to set up the structures of constraint that allow meaning and agency to emerge at all. These can be contextual constraints; the narrative logic; the norms of social interaction; the rules of any game. The author has to provide these constraints so that a user watching or interacting with the spectacle can decode and interpret events and act. It is ridiculous to to try to abrogate the author's responsibility for exacting a specific performance of agency from the user – that's what she's paid for. Within this framework, interaction works by trapping the user, so that her actions implicate her even more strongly in the ongoing semiotic process. The idea of creating a trap in fiction is an old one; Barthes uses the term "snare" to talk about the process of setting the user's

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 125–128, 2005.

mind to work along a certain, desired, path of signification [3] p32. I have co-opted this term for the dramatic structure that has been emerging from our interactive drama practice [1,2].

2 The Play's the Thing Wherein I'll Catch the Conscience of the King

Interactive fiction should be a series of snares for the user; snares that manipulate the user's actions by manipulating the user's emotions [2]. Providing a series of self-referential scenes that slowly reveal information and surprises, forming emotional highs and lows and raising tension, is a staple of drama. This process can be seen as luring the user along an emotional path, a psychological journey. If the journey is interactive, it is not enough to set up the conditions that will evoke the emotions, but we must provide actions that suit that emotional scenario. Providing actions is also not enough, but the actions must be specifically designed to reveal the state of mind of the user. The dramatic progression and success of the interactive drama depend on a process of feedback and interpretation of the user, so that the world and the characters in the world are appropriately responsive. The snare must be baited by the author, activated by the user, and checked by the author; and the results then used to influence future snares. In the following sections I use examples from my own work, The Thing Growing, a virtual reality drama for immersive CAVE-like VR systems[1].

2.1 Baiting the Snare

The bait is the emotional stimulus plus a possible action(s) dangled in front of the user, inviting her to the next step of performed agency. The author sets up a constraining narrative context and/or using intelligent agents establishes a constraining social context. For example in my immersive VR interactive fiction, The Thing Growing, a social/emotional context is established in which a creature the user releases from a box, proclaims its love for the user, and invites her to dance. Common sense psychology tell us that humans are very apt to physically mimic the movements of other humans, that dancing signifies intimacy, that some people are awkward with their bodies. The snare is set.

2.2 Activating the Snare

The action or actions must be deliberately designed to fit in seamlessly and logically with the narrative and emotional context; to be checkable; to reveal as much as possible about the user. The user/agent must understand the signs accurately, and allow her emotions to be stimulated. She must decode the affordances, assess her options, and commit herself to the performance of an action that follows from her emotional state. Continuing the example above, the user understands the invitation to dance and reacts emotionally. She must also understand how she can act - in the case of this immersive VR project she must understand that her body is tracked and she is expected to use it. Her action will stem from her emotion, but a variety of emotions may be stimulated. Some users, rather liking the chirpy creature who is in front of them, perhaps flattered by the declaration of love, feel happy and co-operative. They dance. Other users may like the creature but dislike dancing, they do not. Still others, may already have decided to dislike or distrust the creature. They also refuse to dance.

2.3 Checking the Snare

The snare must be designed so that when we check it we learn as much as possible about the user. Checking the snare depends first on hardware. In the case of the VR example I am using, the user is attached to a tracking system, and has a 3D joystick to move around the virtual environment. Data from the tracking system can be used to assess the movement of the user’s hands and head. Data from the joystick can be used to assess the user’s position in the world and relative to the creature she is with. We backchain from the data we receive to the action that has caused it, to the probable emotional state of the user. So, if the tracking sensors move in a particular way, we understand that the user is dancing, we interpret that to mean that the user likes the creature, is happy and co-operative. Other feedback data indicates the user has turned away from the creature without dancing. Either the user doesn’t like dancing or doesn’t like the creature.

2.4 Building Drama with Snares

One snare is not likely to make interesting drama. For interactive fiction, snares only become meaningful when they are combined. The knowledge gained about the user from one snare is used in selecting others, so that the system responds to the user, maybe surprising her by its response, heading her off, herding her to the next snare, the next check point. This does not have to be extremely complex a simple feedback loop of snare and check can be established.

Returning to our example: the creature responds to a user who is inclined to dance by praising her, teaching her a new dance step, and again checking the user’s action. Dancing obediently the user learns to expect praise - thus a new snare is set - one specifically for this compliant user. Because over time the creature becomes pickier, criticizing the user’s performance. The new snare relies for its emotional impact on two different assumptions about how people work psychologically. First there is a tendency for people to become self-conscious physically if they are criticized, which is heightened because the user is expecting praise. Second some people can be hooked into a psycho-dynamic of trying to please the other in a relationship, to seek out affirmation from the other. In either case a combination of snares leads the user to become unbalanced, unsure, maybe anxious to please, maybe annoyed. If the user continues through several check points trying to dance we assume the former, if they stop dancing we assume the latter. Further snares await her response.

Although the snare is designed to evoke an emotional response which should lead to an action, the author cannot expect to be able to cajole the same performance of agency out of each user. The authorial system must be able to handle a variety of responses. However, our experience building this kind of supported improvised drama with snare-like structures suggests that users tend to fall into patterns of response. Therefore, a strategy of anticipating responses using a common sense understanding of psychology to predict them, and also iteratively testing and refining the snares in order to add necessary responses can result in a drama that works – for some of the people some of the time. Others will not consent to the performance of agency that the piece demands, they will not co-operate in the creation of meaning, and there will be no fictional experience.

3 Conclusion

Each person brings their own instincts, background, and assumptions to a new enterprise such as interactive fiction. My background in experimental fiction and video art has focused on questions of the formation of identity. A first person interactive form is an important new medium for authors, and and exciting playground for participants, in which to explore such conceptual territory. The Thing Growing's was entirely driven by the content we wanted to present, the experiment was to devise methods for delivering that specific interactive experience. Now we are in the process of trying to analyze, formalize and generalize methods that worked to evoke the kind of user response we wanted, and apply them to our new work The Trial The Trail. The elaboration of the snare structure is the core of this work.

References

1. J. Anstey, D. Pape, and D. Sandin. The Thing Growing: Autonomous characters in virtual reality interactive fiction. In *IEEE Virtual Reality 2000*. IEEE, 2000.
2. J. Anstey, D. Pape, S. C. Shapiro, O. Telhan, and T. D. Nayak. Psycho-drama in vr. In *The Fourth Conference on Computation Semiotics*, pages 5–13, 2004.
3. R. Barthes. *S/Z*. Hill and Wang, New York, New York, 1974.
4. T. Kjolner and N. Lehmann. Uses of theatre as model:discussing computers as theatre - some additional perspectives. In L. Qvortrup, editor, *Virtual Interactions:Interaction in Virtual Inhabited 3D Worlds*, pages 76– 93. Springer-Verlag, 2001.
5. B. Laurel. *Computers as Theater*. Addison-Wesley, 1993.
6. J. Murray. *Hamlet on the Holodeck, The Future of Narrative in CyberSpace*. Simon and Schuster, New York, New York, 1997.
7. S. Nath. Story, plot & character action: Narrative experience as an emotional braid. In *Technologies for Interactive Digital Storytelling and Entertainment Conference*, pages 1–18. Fraunhofer IRB Verlag, 2003.

Environment Expression: Telling Stories Through Cameras, Lights and Music

Celso de Melo and Ana Paiva

IST- Technical University of Lisbon and INESC-ID,
Avenida Prof. Cavaco Silva – Taguspark,
2780-990 Porto Salvo, Portugal
{celso.de.melo, ana.paiva}@inesc-id.pt

Abstract. This work proposes an integrated model – the *environment expression model* – which supports storytelling through three channels: cinematography, illumination and music. Stories are modeled as a set of *points of interest* which can be characters, dialogues or sceneries. At each instant, audience's focus is drawn to the highest priority point of interest. Expression channels reflect the type and emotional state of this point of interest. A study, using a cartoon-like application, was also conducted to evaluate the model. Results were inconclusive regarding influence on story interpretation but, succeeded in showing preference for stories told with environment expression.

1 Introduction

Digital technology progress inspires replication of the kind of storytelling seen in the arts. In theatre, a story is told using drama, sceneries, makeup, lights and sound. In cinema, the camera introduces a new expression channel. In digital worlds, besides similar channels, synthetic characters and virtual environments bring new potential.

This work is about the role of virtual environments in virtual storytelling. In concrete, an integrated model is proposed which supports storytelling through three channels: cinematography, illumination and music. Each channel reflects differently, at each instant, the character, dialogue or scenery which is being given focus.

Regarding the rest of the paper, section 2 describes the model and channels, section 3 presents an evaluation of the model and, finally, section 4 draws some conclusions.

2 Environment Expression

Stories, which can be scripted or interactive, are modelled as a set of *points of interest* which compete for the audience's attention. These can be of three kinds: (1) *characters*; (2) *dialogues*; (3) *sceneries*, i.e., the story's setting itself. Each point of interest is associated with a priority which can change through the course of the story.

The *environment expression model* receives as input a story. The story is told through the director and three expression channels. The *director* focuses, at each instant, the audience's attention to the highest priority point of interest. Then, the

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 129 – 132, 2005.

three *environment expression channels* – *cinematography, illumination* and *music* – present it to the audience's senses. Fig.1 summarizes this model.

Emotion synthesis and expression are also explored in this work. Regarding synthesis, a character has an *emotional state* based on the OCC emotion theory [1]. Dialogues' and sceneries' emotional states reflect the participants' emotional states. Regarding expression, emotion is expressed through the expression channels. More information on this work's emotion synthesis and expression is found in [2].

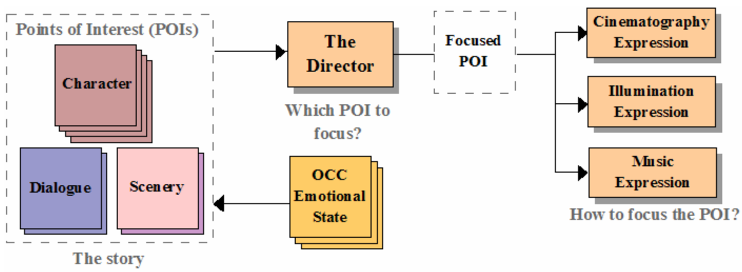


Fig. 1. The environment expression model

2.1 Cinematography

According to [3], camera shots vary according to *distance* – extreme close up, close up, medium, full shot and wide shots - and *movement* – point of view and establishment shots. Distance influences the audience's emotional attachment to the point of interest. Point of view shots convey the point of interest's perspective. Establishment shots overview the setting thus, contextualizing the action. Regarding shot sequences, it is important not to cross the line of action to avoid disorienting the audience. In this sense, for dialogues, the *triangle principle* is widely used.

In this work, the cinematography channel selects the shots according to the focused point of interest's type as follows: (1) if it is a character, a full shot or point of view shot is randomly chosen; (2) if it is a dialogue, the appropriate triangle principle shot is applied according to who is talking; (3) finally, if it is a scenery, establishment shots are chosen. To avoid confusing the audience, camera shots can change only after a certain amount of time has elapsed. Finally, shot distance and angle vary according to the focused point of interest's emotional state [2].

2.2 Illumination

According to [4], the *three-point lighting* technique, widely used to focus characters, is composed of the following lights: (1) *key light*, which is the main source of light; (2) *fill light*, which is a low-intensity light that fills an area that is otherwise too dark; (3) *back light*, which is used to separate the point of interest from its background.

In this work, the illumination channel uses a variation of the three-point technique to illuminate the focused point of interest. The key light is a point light placed between the point of interest and the camera along the 'point of interest-camera' vector. Because it has attenuation and it need not focus a specific target, a point light was preferred to directional and spotlights. Key light attenuation and color varies according to the point of interest's emotional state [2]. The fill light role is assumed by general scene illumination, responsible for illuminating the rest of the scene, and back lights were ignored for their low visual impact and performance reasons.

2.3 Music

According to [5], the relation between music and emotion varies according to: *Structural features* – which relate the music's structure with emotions; *Performance features* – which refer to the influence of the *interpretation* of the music. Regarding the first, tempo is one of the most influencing factors affecting emotional expression in music. Fast tempo may be associated with happy/exciting emotions and slow tempo with sad/calmness emotions. Regarding performance features, the expressive intention of the performer is converted into various cues during the performance.

In this work, the music expression channel reflects the focused point of interest's *mood valence* – positive, neutral and negative. To convey mood valence, music, with the same valence, is randomly selected from a library. To fill the library music was selected according to the following simple criteria: (1) Positive songs should have fast tempo and, if applicable, should have positive lyrics; (2) Neutral songs should have medium tempo; (3) Sad songs should have slow tempo and, if applicable, should have negative lyrics. Regarding the association between lyrics' and music's valence, as the performer conveys the music's mood through cues, it is reasonable to expect that the lyrics' mood propagates to the performance's structural features.

3 Study

To test the model the *dancing solids* application was developed. This is a cartoon-like application with a simple story: "Once upon a time, there was a bunch of dancing solids. There were pyramids, cylinders and ellipsoids. There were boys and there were girls. The girls allured the boys. If the boy liked the girl, he would court her with a dance. If they both liked each other they simply got married. The end". The cartoon paradigm was chosen because it facilitates emotion expression which is essential to the story's plot. Furthermore, to force the audience to rely on environment expression for interpretation, characters were given simple bodies – geometric solids with eyes.

A study was conducted on this application to assess the relevance of environment expression on the audience's interpretation of the story plot, as well as the adequacy of this work's approach for each expression channel. The study, which was fully automated, was organized into four parts: (1) *Subject Profile* – where the subject's profile was assessed; (2) *Emotion Perception* - where the subject was asked to guess which emotions were being expressed by dancing solids using varying configurations of the expression channels; (3) *Music emotional valence* – where the subject was asked to classify 12 music compositions according to one of the following mood valences: positive/happy; neutral; negative/sad; (4) *Stories interpretation* – where the subject was told it would see two different stories when, in fact, it was presented with two different versions of the *same* dancing solids story. Stories were assigned randomly a happy – girl marries boy – or unhappy ending – girl doesn't marry boy. Version A had no environment expression, while version B had all three channels active. After visualizing each version, the subject was asked whether the girl liked the boy. Additionally, before finalizing, it was asked which was the preferred version.

The study was presented to 50 students at Technical University of Lisbon. Subject average age was 23 years and 84% were males. Average time per inquiry was 12 minutes. Regarding emotion perception, results showed that environment expression, in particular the illumination channel, positively influenced emotion perception [2]. Regarding music emotion valence, average subject classification matched predictions for 92% of the music thus, revealing that mood valenced music selection based on tempo and lyrics emotional valence is sufficient to produce satisfactory results. Regarding stories interpretation, in general, subjects perceived correctly the stories' ending independently of environment expression (94.3% for version A and 82.9% for version B). This doesn't mean that environment expression has no influence on interpretation but, that the dancing solids, even though having simple bodies, were expressive enough to tell the whole story. Finally, regarding story preference, when the ending was unhappy both versions were equally enjoyed (60% of the subjects) followed by version B (35%). When the ending was happy, version B was preferred (50% of the subjects) followed by version A (33%). Thus, it seems that even though the study didn't succeed in showing the influence on story interpretation, it succeeded in showing environment expression relevance for the whole storytelling experience.

4 Conclusions

This works proposes a model for storytelling through three channels: cinematography, illumination and music. Stories are modeled as prioritized points of interest which can be characters, dialogues or sceneries. During storytelling, at each instant, the highest priority point of interest is focused differently by each of the expression channels.

Evaluation of this work revealed that: mood valenced music selection based on tempo and lyrics emotional valence is sufficient to produce satisfactory results; subjects perceived the story independently of environment expression; subjects preferred a version of a story told with environment expression to one which did not.

References

1. Ortony, A., Clore, G., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press (1988)
2. de Melo, C., Paiva, A.: Environment Expression: Expressing Emotions through Cameras, Lights and Music. Submitted to the 1st Affective Computing Conference (ACII05) to be held in Beijing, China
3. Arijon, D.: Grammar of Film Language. Silman-James Press (1976)
4. Birn, J.: [digital] Lighting & Rendering. New Riders (2000)
5. Juslin, P., Sloboda, J.: Music and Emotion: theory and research. Oxford University Press (2001)

Interactive Digital Storytelling

Toward Interactive Narrative

Ken Perlin

Media Research Laboratory,
Department of Computer Science,
Courant Institute of Mathematical Sciences,
New York University, USA
http://mrl.nyu.edu/~perlin/

What is the future of interactive entertainment? Can we tap deeper emotions? Can we go beyond game-like experiences to create powerful interactive literary narratives? Are these even meaningful questions? In this paper I will address one possible way to make these questions meaningful, and one possible path to their answer.

First let's briefly discuss the significance of *stories* within culture. Stories are actually quite central to cultures. One could even say that a culture defines itself by the stories it tells.

But what if we want to interact with stories? This would raise several questions: *(i)* what would change? *(ii)* what would stay the same? *(iii)* how do we make such a thing? and finally, *(iv)* where is the artist/author then located, with respect to the observer/reader?

One way to approach these questions is to try to look at the act of reading or hearing a story as a peculiar form of game-play. To do this, we need to examine why stories are able to be so compelling.

A first observation is that the mainstream obsession of human culture is other people. We are *all* continually engaged by questions on the order of:

- "Why did she just say that?"
- "I wonder what they were thinking?"
- "Do you think he really likes her?"

These are not scientific questions. Rather, they call for shaded emotional judgments and intuitive understandings of human interaction, as well as the subliminal ability to "read" people. Such questions exist within a space of psychological awareness.

To relate this space to interaction, it is useful to compare our traditional notions of game play to our traditional notions of narrative.

Narrative is generally driven by character. The unities that generally bound the telling of narratives are quite strong, even when those narratives appear in widely varying sensory forms, such as novels, plays, movies, or short stories. Consider, for example, the Salinger novel "Catcher in the Rye" and the MGM film "Gone with the Wind" (not the novel):

On the sensory level, Holden Caulfield is experiences only as printed text on paper, whereas Rhett Butler and Scarlett O'Hara are experienced via the appearance and

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 135 – 147, 2005.

speech of skilled actors projected onto celluloid. Yet our *reasons* for following their stories are quite similar. Each character has a central character mystery, some set of desires and character traits that are gradually revealed as the narrative progresses. We turn the page or keep watching the screen mainly to find out what these particular powerful personalities will do next as the plot unfolds: how they will respond; what choices they will make.

In fact we can state that linear psychological narratives generally have in common *(i)* psychological buy-in by the audience and *(ii)* a requirement of willing suspension of disbelief.

The reader/viewer is continually being challenged by such questions as: Who are these people? What will they do next, and why?

It is not *what* happens that primarily holds our interest, it is how we believe the characters feel about it. Note that the reader/viewer has no agency over the plot or its characters. The only "action" that the audience takes is an ever-changing internal emotional empathic response as the characters experience and respond to their world.

One of my favourite sayings about the centrality of character in stories is the following: "Plot is the drugged meat that you throw over the fence to put the dog to sleep, so you can rob the house."

Contrast this with games. In a game you have quite a bit of agency over what happens, through some sort of game mechanic (e.g.: moving your avatar forward, jumping, shooting, opening doors, solving puzzles, picking up objects, ...) and yet you generally don't believe in the characters in the deep way that you believe in the characters of a novel or film. When I play The Sims, everything in the game is telling me that the people I see are doll figures (albeit very entertaining ones), not deeply resonant characters like Holden Caulfield or Scarlett O'Hara.

And yet, we can actually cast the experience of reading a novel or viewing a movie as a peculiar sort of game mechanic. This game mechanic is one that I call "hack the characters". As the story progresses, we continually find ourselves asking questions, such as: *"who are they?" "why are they doing that?" "what will they do next?"* In an important sense, we *know* that we are playing a sort of game, and that it is this peculiar game play that makes watching a movie quite different from experiences in real life.

Take for example, the following scenario. A husband and wife are hanging around their kitchen on a relaxed Sunday afternoon, when the husband's best friend walks in. The wife immediately becomes quiet, and pointed doesn't look at or talk to the friend. The husband doesn't notice.

If this were to happen in real life, there could be a million different underlying causes for these events. But if it happens in a movie, the audience immediately knows that the wife and the best friend are sleeping together behind the husband's back, and the audience is engaged in the underlying question of if or when the husband will find out.

One can describe many such scenarios. As our response to these scenarios makes clear, movies and other stories are full of artificial conventions. We are all so used to these conventions that we are often not even consciously aware of them. This graceful slipping into stylized convention is essential to the willing suspension of disbelief that allows the narrative to reach us and to exert power over us.

More specifically, when you pick up a novel or start to watch a movie, you know at some level that you are engaging in a contract with an author. There is a *reason* the author is leading you through this story, and that reason generally centers on creating within you an emotional bond with the characters in that story, and in giving you, by proxy, an emotionally dynamic experience through those characters and through the choices they make.

This is why psychological narrative is interesting, even though the reader or viewer has no control over what happens. Even though *you* can't change anything, you are entertained by the *characters'* choices. This process requires the author to carefully maintain the believability of characters. If Rhett Butler were to suddenly dance the hula with a chicken on his head, Gone with the Wind would lose a considerable amount of its hold on the audience.

If we wish to make interactive psychological narratives that allow their audience to engage in a similar process of "hack the character", we need to respect this principle of believability.

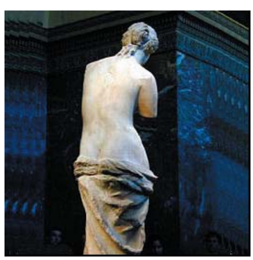
For example, if you were to make an interactive version of *Raiders of the Lost Ark*, you would not want to allow the player to fly the camera into Indiana Jones' underwear drawer to find out whether he wears boxers or briefs. Such an interaction, being on a completely inappropriate level, would destroy the essential mystery of Indiana Jones, and would completely pull you out of the interesting questions about his character that make him deep and compelling.

Similarly, you could not effectively engage in a process of "hack the character" if you were allowed to hack directly into his sub-conscious and tweak the knobs there, which is effectively the game mechanic of The Sims:

We can state such limitations on game-play as a principle: The player should only be able to interact with or influence things *outside of* characters, as those characters make their choices.

This whole idea of defining a genre of art or entertainment by defining its scope of interactivity is actually quite old. Take, for example, the difference between the traditional genres of painting and sculpture. Even if we look at a painting such as Leonardo DaVinci's Mona Lisa in different ways, we essentially get the same viewing experience:

In contrast, viewing a sculpture such as the Venus de Milo is a different and unique experience for every variation of view angle and lighting:



It is important to note that this differing and unique experience for each viewer does not imply that each visitor to the Louvre is an artistic collaborator. It simply means that sculpture is one of many art forms that are meant to be experienced in an infinite variety of ways, with a different experience available for each recipient of the work. Other examples of this are architecture, musical instruments, and the creation of procedural art with software.

So what might be an interactive “story-like” game mechanic that preserves character believability? An example will help here: a bar scene (bar scenes are always fraught with possibility).

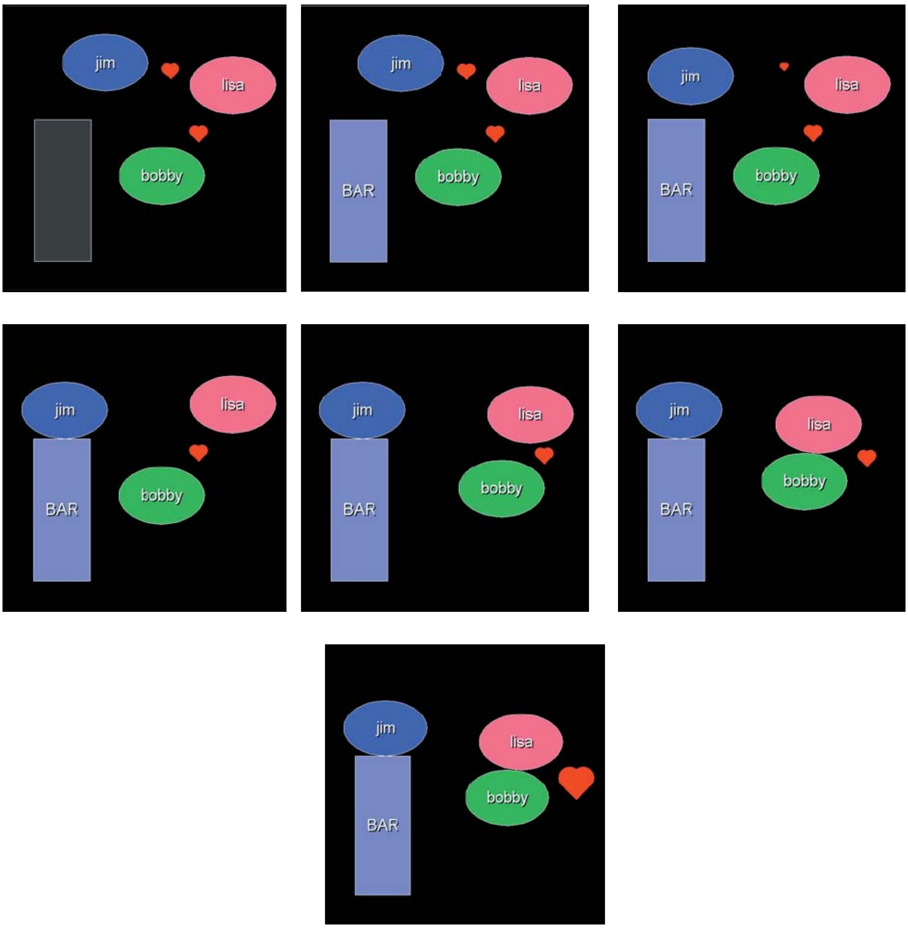
Let us say that our young protagonist Lisa is interested in two young men: Jim and Bobby. Jim is handsome, dresses well, is quick with a compliment or clever line, and is popular with the girls. Bobby, in contrast, is less good looking, dresses unfashionably, and isn’t nearly as quick or clever. But we already know enough at this point in the narrative to know that if Lisa were to end up with Jim tonight, he would just sleep with her and quickly go on to another conquest. She would likely become depressed, go off her diet, drop out of Medical School, and generally become a less interesting character.

Now at this point is important to keep in mind that *if* we maintain believability, then interesting characters are always more rewarding for the player than uninteresting characters. They player *wants* Lisa to stay interesting and full of possibilities.

Note that if the player were able to directly hack into Lisa’s subconscious, as in The Sims, then it would be impossible to keep Lisa interesting in the traditional narrative sense, because Lisa could then be readily nudged into betraying or subverting her own goals. It is this uncanny sense of unbelievability that moves players of The Sims to set fire to their characters, drown them, starve them, and otherwise wreak havoc upon them the way that some small children experiment with ants and magnifying glasses on a sunny day (all of which was quite anticipated by the game’s designer Will Wright, incidentally). In contrast, a reader of Catcher in the Rye would never be moved to do such things to Holden Caulfield. He is much too valuable alive, as a continually surprising character - the reader very much wants to see what he is going to do or say next.

So what sorts of game mechanic are available to the player, outside of modifying the character herself? I posit that the proper way to maintain believability is to provide the player with a constrained ability to modify the world *around* the protagonists, and to follow the response of the protagonists to these modifications of their world.

For example, let's say that we see Lisa in the bar with Jim and Bobby, and we happen to already know that Jim has a weakness for drink. We can't *make* him do anything, but we can tempt him. Let's say the player can "influence" the world to the limited extent of opening the bar, at a point in the conversation where Lisa is deciding between Jim and Bobby. The sequence below shows how this might play out: Jim moves to the now-open bar, thereby focusing less of his attention on Lisa. Lisa therefore has a chance to spend some time with Bobby, and the balance of power shifts. Lisa and Bobby bond, she ends up with him, stays on her diet, finishes Medical School, and is ready to move on to ever new and exciting adventures.



But none of this works if the world around the characters is not believable. To go back to an earlier example, if the player can suddenly place a chicken atop Lisa's head, the entire question of psychological believability becomes meaningless. In order to keep the whole enterprise from falling apart, we need some reasonable - yet non-intrusive - way to constrain the player's choices.

One way to do this is by thinking in terms of parallel universes. As time goes on, possible universes appear at a more or less constant rate. Every time the player makes a choice or decision, this forest of potential universes is pruned. This sort of pruning was nicely illustrated in the 1998 Peter Howitt film "Sliding Doors". In that film, the story split into two possible futures: one in which the heroine just barely caught a train, and the other in which she just barely missed it. Note that if this had been an interactive game, rather than a movie, and the player had chosen one of these futures, then in that moment, the number of possible outcomes for the game would have been halved.

Such a situation arises every time a player makes a choice in a game. In this sense, there is always some sort of equilibrium: As time moves forward, more possible universes appear and multiply. Meanwhile, as the player makes choices, possible universes disappear and divide. A world can be said to be "believable" if this equilibrium is reasonable. A world can be said to be "unbelievable" to the extent that the player is making choices that divide the character's possibilities so severely as to make them uninteresting, such as by forcing Lisa to walk around with that chicken on her head.

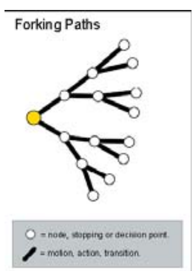
We can even formalize this a bit. If one equal choice by the player places the narrative into a future having probability 1/2, then two successive equal choices by the player will place the narrative into a future having probability 1/4. And in general, *n* successive equal choices will place the narrative into a future having probability $1/2^N$. For example, twenty successive even choices would lead to a future having probability of about 1/1000000.

To enforce believability, we can maintain some sort of cost for making choices. For example, the player can be given a certain store of spendable energy. Making a choice costs a certain constant amount of this energy. This leads to the following property:

$$\text{Energy} \equiv -\text{Log}(\text{probability})$$

So how do we author systems that allow artists to create a mutable narrative that can evolve in a way that reflects user choices?

We cannot use explicit branching narrative structures, because this leads to a combinatorial explosion:

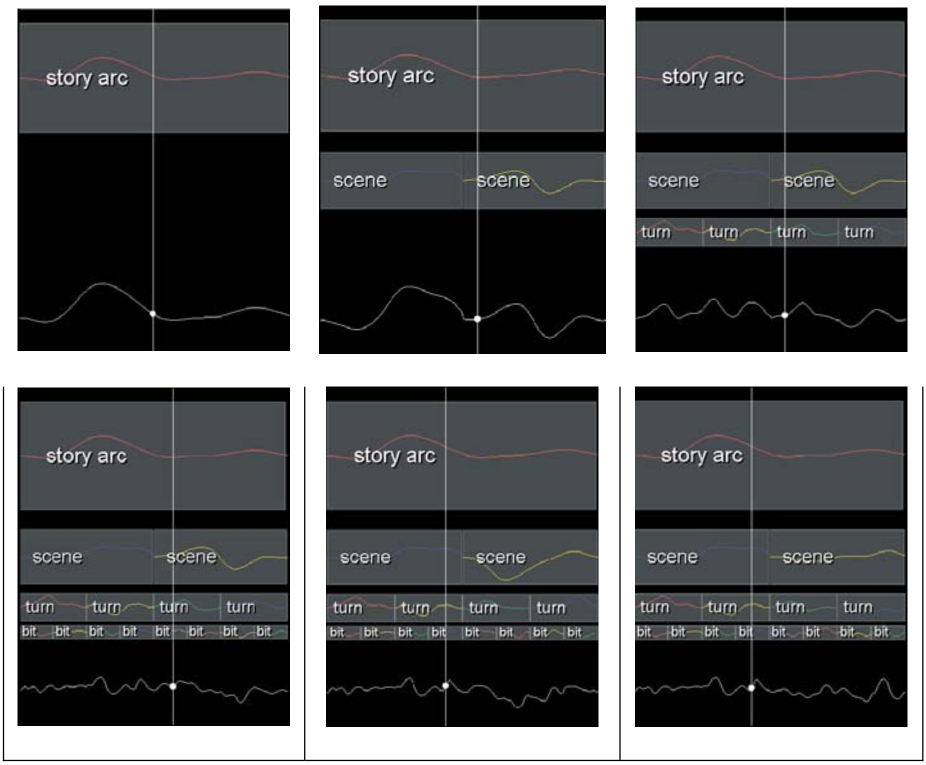


The amount of work it take to make anything interesting through the creation of branching narratives is exponential: In order to support a typical N-decision path through the narrative space, the author needs to have authored on the order of 2^N distinct story segments.

What is called for is multi-layered interaction. The author needs to be given tools to define desired properties of the interactive narrative at a coarse-grained level, as well as tools to define properties at successively finer levels of detail.

The response to an audience intervention must be mediated at all levels of detail. If the audience has just made such an intervention (such as opening the bar in the example above), then the question "what happens now" is not a linear one. Rather it involves the contribution or style sheets at multiple levels of detail.

We can define this as the principle of layered contingent narrative. An engine that supports believable interactive narrative properly looks like a fractal, with influences and interreflections between different levels of detail.



Using Style-Sheets When Authoring Interactive Media

Consider the user interaction paradigm of a Web page built with some commercial layout package such as Front Page:

```
user text
• { layout software • style sheet }
• { renderer • fonts/kerning }
• document view
```

There are four different types of people involved in the user of such a WYSIWYG editor: *(i)* the author who is creating the content, *(ii)* the user who is browsing the page, *(iii)* a page layout expert who is making decisions about spacing, centering, etc., and *(iv)* a programmer who implements that connection between the first three people.

Generally speaking, the author of the document never actually needs to meet the stylist or the programmer. As far as she is concerned, there could be an entire team of page layout experts and programmers somewhere behind the scenes (and there generally is). Some of these people may even be deceased by the time the web content itself is written.

Yet the talent and ability of the page layout expert is available to the author of the interactive document, as long as the expert has encapsulated that talent and ability into style sheets which can be interpreted by the program and applied to the author's document.

The key here is that creation has been divided into two processes, with two different languages: *(i)* a style description language, for specifying geometric relationships and behaviors for entities on the page, and *(ii)* a content language, which is mostly a natural language such as English.

Note that there must be another programmer/style-expert pair underneath the first one, simply to deal with decisions concerning character shape and font kerning. So we can see that the principle of content creation supported by programmers and style experts extends to layers of support. Every time the author adds, deletes or rearranged words, a team of experts is virtually in the room, making corresponding spatial layout decisions based on generic (ie: not specific to this specific content) principles of page layout design.

Principles for a Narrative Generation Architecture

An interactive narrative that is played out in a software run-time engine can similarly be designed as an interaction between author, audience, and visual style expert, where the latter's contribution is supported by software layers that convert style advice into run-time actions:

```
A user's contingent script:

•{ behavior system • behavior script }
         (produces discrete tokens of action and mood)
• { animation system • animation script }
         (produces joint movements, voice prosody)
• { renderer • appearance model }
         (produces polygons, sounds)
• animation view
```

In this case there is a programmer/style-expert pair at three different levels: *(i)*what happens next, *(ii)*who such decisions are acted out, and *(iii)*how it all appears visually.

The first (highest) of these levels is concerned with plot decisions, underlying psychological tone, and general scene blocking and camera placement. The second (middle) level is concerned with body language, gesture, facial expression, fine details of both scene blocking and camera placement. The third (bottom) layer is concerned with what we traditionally think of as graphics rendering.

For each of these layers, a combination of talent and expertise is required by a style expert. Also required for each layer is a scripting language built around such talent and expertise, which allows the expert's contribution to be applied to enhance the specific content that was created by the author.

Since such systems are in their infancy, it is difficult to be more specific at this stage. The purpose of this description is to focus future developments toward effective definition and implementation of such style specification languages for interactive narrative and interactive virtual acting.

The ability to author non-linear narrative decisions. is half of the solution to the problem of creating interactive narrative experiences. The other half of the solution is presentation of those decisions to the audience. In a cinematic or game-like medium, this requires acting.

Principle of Believable Virtual Acting

Note that the middle layer of our style diagram is concerned with what is traditionally called acting: how an character turns his head or places her foot; whether a character turns fully to look at another, or only acknowledges the presence of the other peripherally. The ability to author and then play out non-linear narrative decisions requires believable presentation of those decisions to the audience. In a cinematic or game-like medium, this requires acting.

In the absence of believable actors, everything runs the risk of looking like Plan 9 from Outer Space:

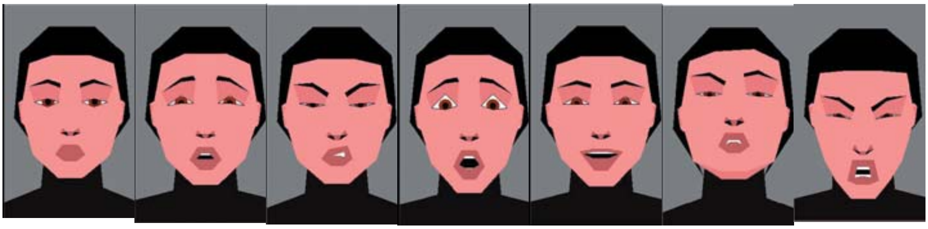
Note that canned linear animation is not sufficient for purposes of implementing interactive narrative. The second time that an audience sees an actor perform the same movements, that actor effectively ceases to exist in the audience's mind, since that character has now been identified by the audience as a mindless automaton, and the all-important sense of "what will the character do next" has been lost.

The following principles need to be adhered to:

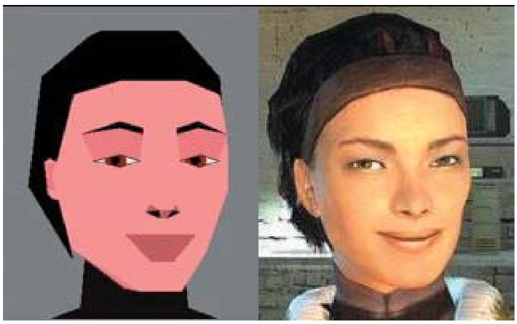
- canned animation is a dead end.
- acting needs to be procedural, from the inside out

Experiments in Believable Virtual Acting

In related work in support of non-linear narrative [4] we have done experiments with emotive facial expression:



Some of this work has made its way into commercial interactive games such as Half Life 2 which does has an underlying narrative, although not the engine needed to maintain character believability. For example, the facial musculature of the character of Alyx in Half Life 2 was influenced by our work on facial expression [1]:



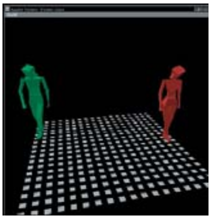
Similarly we have done work on virtual actors that can convey body language, shown briefly below. This work is gradually finding its way into the game industry, but its real utility will come when applied to interactive narrative in support of believable characters.

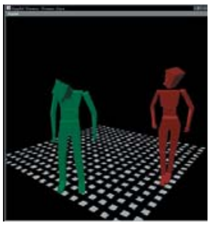
Façade

There have been some attempts to build a framework such as the one we describe. For example, the groundbreaking work Façade by Michael Mateas and Andrew Stern is a working example of contingent narrative [2].

To implement this work they developed the ABL language, which allows contingent execution of narrative plans. For example, if a high level directive of the narrative requires a character to carry a bottle of wine across the room, but the actor already has a cigarette in one hand, then a lower layer of narrative execution directs the actor to carry the bottle with the other hand.

Yet Façade is not quite what we want, going into the future. According to its authors, Façade does not break down the narrative choices in all levels from highest to lowest, but rather relies on several dozen carefully scripted interactive narrative scenelets, with techniques to steer the narrative toward these relatively linear set pieces. The resulting particular interactive narrative had to be heroically built through several years of intensive work. In a sense, one might say that Façade has hit an important half-way point between the traditional forking narrative, and the sort of fully procedural narrative generation system that we have outlined here.

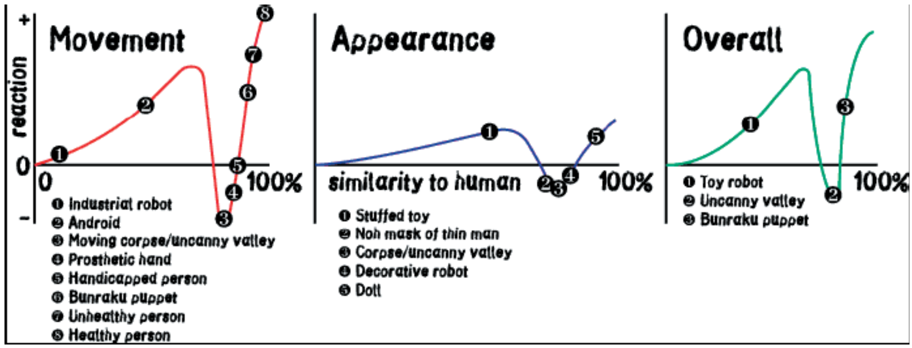


To put the various elements into one place, here are some thoughts about what narrative constraints might be needed in the near term:

- The audience or "player" cannot have an avatar that engages in dialogue with characters, because currently existing techniques are not sufficient to sustain conversational believability between an actual human and non-player characters;
- There can be no direct puppeteering of the minds of characters, in the style of The SIMS;
- The audience cannot freely fly around the camera to peer into the sock drawer of Indiana Jones, or otherwise breaking character believability;
- The audience should only have the power to manipulate the world, not characters themselves; the characters should be free to respond in a way that accords with their nature;
- The *probability* of events happening should be respected.

What About Realism?
Photographic realism is only of benefit for interactive narrative if it can be sustained. Inappropriate realism, brought in without proper support for acting and behavior, actually works against believability. Dr. Masahiro Mori laid out the basis for much of this in his principle of the "Uncanny Valley" - principally in the observation stopping just short of required realism can lead to things becoming very unbelievable [3];



In fact it is often better for characters to remain decisively cartoon-like. A good recent example of a successful application of this principle is Brad Bird's brilliant animated film *The Incredibles*.

Conclusion

In conclusion, we have laid out some principles describing how the tools to create interactive character-based non-linear narrative might be developed, and also what such a medium might look like in practice. If we are collectively successful in implementing such a thing, then we will be able to direct (not just animate) believable interactive actors, and interactive narrative will finally get its Steven Spielberg and its Virginia Woolf. We will attain true interactive character-driven narrative, and our work will touch peoples' souls.

References

1. Birdwell, Ken. 2005. Personal communication
2. Mateas, M. 2005. http://www.interactivestory.net
3. Mori, M. 1982. The Buddha in the Robot. Charles E. Tuttle Co.
4. Perlin, K. 1997. Layered Compositing of Facial Expression. In *ACM SIGGRAPH 97 Visual Proceedings: the Art and interdisciplinary Programs of SIGGRAPH '97* (Los Angeles, California, United States, August 03 - 08, 1997). L. Pocock, R. Hopkins, D. Ebert, and J. Crow, Eds. SIGGRAPH '97. ACM Press, New York, NY, 226-227.

Managing a Non-linear Scenario – A Narrative Evolution

Sandy Louchart[1] and Ruth Aylett[2]

[1] Centre for Virtual Environments, University of Salford, Salford M5 4WT
s.louchart@salford,ac.uk
[2] MACS, Heriot-Watt University, Riccarton, Edinburgh EH10 4AS
ruth@macs.hw.ac.uk

Abstract. This paper examines the causes and consequences of the narrative paradox phenomenon widely observed in VR. We present an alternative approach to virtual and interactive storytelling in the form of the emergent narrative concept, together with an implementation of a subset of these ideas in the FearNot! demonstrator.

1 Introduction

The continuously evolving and developing technology of Virtual Reality (VR) and the rapid technical development of real-time interactive computer graphics has produced a new expressive medium. However, Interactive and Virtual Storytelling (IS/VS) have yet to conclusively address several fundamental and critical issues. What is the role of the user in interactive dramas? What shape and form should they take? How should a story be considered? How should it be articulated in regard to interaction from the user(s) or artificial entities? And how can we manage a non-linear scenario?

In this paper we seek to address these issues by dealing with the *narrative paradox* [1] generated by the conflict between the interactive freedom that is a basic characteristic of the medium [2] and the structure required by traditional conceptions of plot. Indeed, we question the very validity of the Aristotelian plot [3] as a useful principle in the face of interactivity, suggesting that it has dramatically altered the inherent nature of narrative in VR.

We have in previous publications identified the problems posed by the narrative paradox [1], documented the basis for a "bottom-up" approach (i.e Emergent Narrative) to the problem [2, 4], investigated such approach in regard to alternative media practices [5, 6] and applied principles of this approach within the agent-based EU project VICTEC [7]. The work presented in this paper is complementary to previous publications and brings a media oriented consideration of the interactive storytelling concept. It also aims at presenting the authors' particular views on the interactive storytelling concept and lay the basis for future development and theoretical work on the EN concept.

First we present the EN concept as an alternative approach to conventional thinking, and as a potential solution to the narrative paradox. We will then consider the issues arising from the introduction of interactive elements within a narrative frame and its repercussions on the way interactive narratives might be perceived. Finally, we will investigate the potential of the Emergent Narrative approach to

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 148 – 157, 2005.

successfully manage and articulate interactive dramas, citing a small-scale experiment in a constructed system.

2 Emergent Narrative as a Source of Story

Though the merging of emergent and narrative elements as a potential solution to the narrative paradox should be seen as a novel approach, the concepts of emergent structures and bottom-up designs have been studied and developed for many years [8] In particular, these concepts have recently been widely applied to the entertainment industry, in what has been popularly described as “reality TV”. Although not advertised as such, television shows such as “Big Brother ™” in the UK, in which stories emerge directly from the interactions between their different protagonists (i.e. contestants, audience, host), can be recognised as emergent narrative forms.

These programmes, whether we personally enjoy them or not, show that non-linear character-based applications can produce entertaining and meaningful stories. However, one should recognise that these shows are entertaining because they are appropriately edited and that they provide two totally different experiences, depending on whether the user is a participant or a spectator. In this case the spectator perspective is typically more entertaining than that of the participants. In fact, the general feeling inside the confines of the show is often one of boredom, and participants in different series have expressed the desire to leave for that exact reason. The participants are not there to be entertained, they are the entertainers.

In this application of the emergent narrative concept to the television medium, the audience is the prime target and represents the population that should be entertained; the participants are offered the promise of a large sum of money and television exposure to encourage them to play their role in the making of the show. The narrative is managed as a process where participants create the story through their interactions with each other, The role exercised in this case by the programme production team could be related to that carried out by a human Game-Master (GM) in non computer-based Role Playing Games (RPG). We earlier identified [5] that one of the most important tasks carried out by the GM was to create, with a rich level of detail, the different worlds, props and characters necessary for the game or campaign and their different interactions and relations with each other that would generate potentially interesting situations. The main aim is for these particular interactions or pre-determined events, is to provoke dramatic reactions within the players. This role is indeed perfectly assumed by the programme’s production team.

In any character-based drama or application, the choice and definition of the main protagonists, the contestants in this case, is essential to the generation of interesting and potentially dramatic events and ultimately stories or micro-stories. The contestants are selected with great care in relation to their personalities, beliefs, social backgrounds, mental strength or sexual orientation. By selecting contestants with conflicting characteristics, the production team increases the chances of interactions between them creating dramatic events and potentially interesting narrative elements. In most cases such interactions arise from emotions such as tension, frustration, passion or temper, resulting in diverse emergent conflicts and alliances.

The world environment itself is also carefully designed to foster these emotions and deny the contestants any possibility of escaping the constant pressure surrounding them, for example through rules preventing contestants from entertaining themselves by any means other than interacting with each other (i.e. writing, reading). Although, the situations generated in the show emerge from direct interactions between contestants and could not possibly have been planned in detail beforehand, certain likely or recurrent scenarios are taken into account. For instance, it makes perfect sense to design a remote area where contestants can experience relative privacy, as long as there are enough cameras and microphones in that area to cover every possible angle. This allows for interactions that are generally seen as private or secretive such as conspiracy, passion or private arguments.

While design is prior to the performance itself, there are also elements of performance-time control. These are limited to the introduction of pre-determined events that establish or create situations likely to generate interesting actions or reactions from the participants. For instance a daily task in which one half of the participants is given food and the other starved for a day creates a certain tension in the group resulting in a higher chance of dramatic interactions. Thus the production team, acting as GM, manages the intensity and the general mood or feeling of the performance from a character perspective to produce emotionally engaging situations for the audience. However, this format does not suffice in itself to provide an entertaining programme that would trigger audience interest sufficiently for them to watch the programme on a regular basis. A cleverly edited footage is thus presented regularly summarising the activity of the contestants and reporting in a dramatic fashion on their feelings, moods, outbursts or attitudes towards each other.

Although such programmes are good applications of the emergent narrative approach, they embody only one form of the concept applied to one particular medium (i.e. television). Moreover since the target is not actively participating in the show, it does not bring any direct solution to the narrative paradox. However, there is much to learn from this particular type of EN and its study can help in shaping a suitable form adapted to immersive interactive media. Several major alterations must be implemented if one is to adapt the EN theory to VR in order to move from a spectating form to the performative one appropriate to an immersed user.[**Table 1**].

The role and nature of the user is crucial for two somewhat different forms of emergent narrative. If the user is part of an audience and does not have any active role within the performance, the focus of attention will be on providing user entertainment, even if this is to the detriment of the participant's experience. The "spectating" form of EN aims at providing something interesting to watch, look at or read. This is not dissimilar to organic improvisational drama, where the director, in rehearsals, creates certain situations for actors to improvise. The best and most interesting interactions will be then used as raw creative material for a theatrical piece.

The "performative" form of the EN approach is different in nature. Since the main centre of attention is a participating and active user, the focus for the performance is to give the user an enjoyable and compelling experience. Thus, the aim is to produce something that is interesting to live and experience rather than watch. Indeed, the whole experience need not produce anything visually presentable or particularly interesting from an external perspective as long it was when lived and performed from an internal perspective.

Table 1. Performative and Spectating Emergent Narrative forms

Emergent Narrative forms		
	Spectating form	**Performative form**
Medium	Television	Virtual Reality
Target audience	General public	User(s)
User role	Inactive (mainly)	Active - participating
Emergent source	Contestants	User(s) and virtual agents
Time constraint	1 hour summary	Real-time
Cast	Rich conflicting characters	Rich conflicting characters
Managing constraints	Public interest Intensity of the show Contestants, moods, emotions and feelings	User interest Feeling of immersion Intensity of the performance User mood, emotions and feeling
Drama manager	Authoritative and visible	Distributed and hidden (in agents)
Dramatic tools	Narrative and situational events, editing	Narrative and situational events

3 Re-evaluating Narrative Concepts for VR

We have previously highlighted the need to approach the role of the user according to the nature of the performance or the medium [9]. The shift of attention from spectator to "spect-actor" (i.e. Augusto Boal's terminology – [10]) or participant in an interactive scenario focuses on interactivity and immersion. However, such shift in emphasis cannot be made without a major rethink on the exact nature of stories, their requirements and their definitions. The narrative medium has evolved with the introduction of interactive features or elements - particularly true for VR – while the set of narrative tools at our disposition to analyse, understand and generate such interactive narratives has not been developed with respect to this particular characteristic. Although several landmark works [10, 11] have been published over the years, work in the domains of IS and VS is still at an experimental stage. Arguably, interactivity represents the most significant technical change in narrative terms since the invention of cinema and requires a new examination of the basic concepts.

3.1 Story

The first such concept is that of story itself. How can a story be depicted and executed so as to include interactivity with a user in a virtual environment? A story-line or plot exercises constrains the user's freedom of movement and action within the virtual world, while allowing the user interactive freedom affects the unfolding of the story, hence the narrative paradox.

It appears that once interactivity is involved, story becomes plural. All of the different approaches studied in recent years, branching [12, 13, 14] or emergent, actually deal with multiple stories. In the case of branching systems, the stories potentially displayed and executed are all instances or variations of a given story, while in emergent concepts, they result from the association of many micro-stories at character level. Although multiple storylines are common in literature, cinema or even theatre, VR presents the major characteristic that changes or alterations made in these sub-stories are orchestrated consciously or not by the user. This makes the execution of an Aristotelian plot a difficult challenge, in terms of timing and outcome from a branching point of view and in terms of formulation, articulation and representation from an emergent perspective.

Although the general format of beginning middle and end can be respected in principle, an emergent approach to storytelling need not apply it in the sense that it focuses more on the actions and paths of individual characters than on an overall general story. Our definition of what makes a story has to be extended and broadened in the face of interactivity to the consideration of narrative as a dynamic process rather than a static structure. Conventional storytelling (i.e. Aristotelian plot) can then be seen as a support tool that can help in bringing depth, meaning and context to a performance, rather than the directed or targeted objective. To a certain extent, though based on actions rather than character development, this technique is very much in use in the computer games industry, particularly when designing adventure games. The story can then be used for generating interesting and contextually correct events while still not interfering with the user's freedom of movement within the story world.

3.2 Narrative Authoring

Authoring represents a major challenge to this shift in perspective. Authoring, in storytelling terms, is stereotypically the representation of the author's mind, the vision of one person. The author's ability to create interesting stories, characters and narrative events, is coupled with control over the timing, order, rhythm and nature of the different story events and their display. The story, as witnessed by the spectator, is the procession of an appropriately orchestrated narrative vision for dramatic purposes. The idea of an intervening user is incompatible with such an approach since one cannot expect the user to make the right decision for the story, at the right moment or even at the right place, unless the user plays a role similar to that of a conventional actor following a script.

However, if one wants the user to be able to exercise a certain level of freedom of movement within the virtual environment, and, therefore avoid the narrative paradox, one has to consider the user as an autonomous actor rather than an actor in the conventional sense of the term. Autonomous actors would play character parts or roles in a scenario while still enjoying the freedom to make their own decisions at times they feel appropriate. This view of the user would allow both behaviours and narrative events to emerge from interactions and to some extent reconcile narrative input and user freedom.

However this forces a change in the whole concept of narrative authoring; the focus of attention should be on the characters and their ability to interact with each other rather than the overall story. Indeed, the role of the author is confined to writing interesting characters with a strong potential for dramatic interaction together with flexible narrative events whose only aim is to set up scenes and situations. If authors are to write for interactive applications in this way, they must let go some of their authorial powers, a change in the opposite direction to film with its increase in control (i.e. camera angles, music soundtrack etc.). The overall story should be regarded as a hypothetical meta-structure not directly experienced by any character, whose aim is to set up interesting or potentially interesting situations.

It becomes essential for the author to be extremely attentive to the inner state of the user, or in more generalised terms, to the characters' internal states. The primacy of dynamic internal states, and specifically that of the user makes the whole process of evaluating a story difficult and even relatively pointless and puts the focus on the character and its emotional activity. The success of a performance results not from the quality of the story in conventional terms, but the level of enjoyment, active and willing participation from the user.

3.3 Storification

In conventional narrative forms the engagement of the user is reported indirectly by applause or even global sales; in a participative form it is basic to narrative development. Although one can and should analyse signs of enjoyment or immersion of users via their behaviours, level of activity or response within a performance, essential information for the evaluation of such a narrative approach still remains undisclosed and only known to the users. Some can be retrieved through the use of post-performance questionnaires but the subjective story-as-experienced may remain permanently hidden. A feature of live role-play is the debrief at the end, in which the multiple story experiences of the participants are shared and integrated through the appreciation of larger-scale causal chains than those an individual has directly experienced. *Storyfication* [15] is a term that defines the continuous activity of a narrative participant in building a mental picture and developing and testing expectations about the story's outcome and the character's present and future motivations, roles and emotions as the story unfolds in real-time. What separates this process from the variant present in spectating is the situated position of the participant – more limited in terms of global understanding, but richer in terms of ability to act.

In the current absence of non-invasive and reliable mechanisms for estimating user emotional state, one can fall back on monitoring external signs of non commitment, as seen in RPGs [5] where the GM constantly tracks the user's activity or behaviour (i.e. suicidal behaviour, lack of activity, clear lack of interest, lack of attention) in order to assess his/her internal state with respect to the performance. Theatre, cinema and literature have shown that the user's internal emotional state can be manipulated to a certain extent via purposely misguiding hints or indications creating the right frame of mind for a particular effect (i.e. suspense, twist, or surprise), and these techniques need to be reanalysed in the context of EN both for the benefit of authoring and narrative management in real time.

4 Managing an Interactive Drama – An EN Drama-Manager

Story or drama management, is typically where the crunch takes place in interactive narrative. The role of a manager in application of conventional narrative theory is to keep the overall story 'on track' in the face of user actions. Branching narratives avoid the problem by authoring management into the branches. Other approaches involve variants of universal plans [16], whether through the beats of Façade [10] or interacting plan trees [11]. Here management is a navigation issue, but short of anticipating every possible user action in context as in Façade, in which the author faces a combinatorial explosion of possibilities, only certain paths constitute 'the story' and the management dilemma is to how push the user along these paths.

The implication of the arguments advanced so far is that in EN the drama manager should not focus attention on the quality and meaning of the overall story but on the quality of the performance experienced by the different characters (i.e. user, other agents), so that 'staying on track' is no longer an objective. This requires the development of metrics of performance quality, but since it should be measured from the point of view of the different characters, the idea of a distributed story manager within different agents in the world environment is a very natural one.

By equipping characters with an extended action-selection process, in which choice of action is influenced by performance considerations as well as the more usual one of goals and affective state, management would execute below the surface of the visible story and would not disturb the feeling of immersion we are aiming to protect. Global management would then be confined to events exogenous to the characters: entrances, exits, the outcome of unpredictable physical actions (in the absence of comprehensive – and computationally expensive – virtual physics) and, in RPG terms, 'wandering monsters'. Since most of the performance design is directly imputable to the harmonious definition of both the world environment and the characters, as in its RPG counterpart, the role of the drama manager in the EN approach is one of policing the boundaries of character roles and introducing situations and narrative events when required [Figure 1].

This approach has in fact already been the subject of many applications in the domain of RPGs where believable characters are added to a given literary setting and

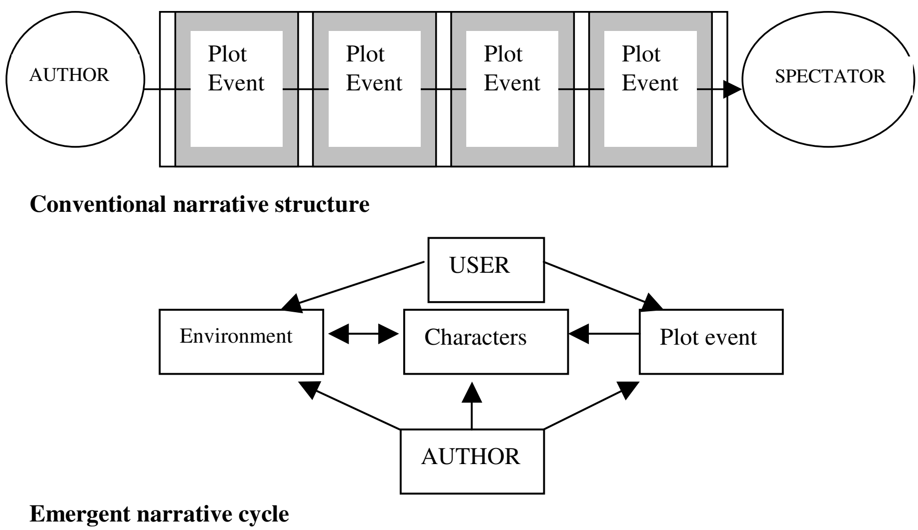


Fig. 1. The emergent narrative articulation

run in parallel to its performance. The pedagogic division of the Gothenburg Region Association of Local Authorities [17] has performed on several occasions a 200-person version of Shakespeare's Hamlet based on this principle where the additional characters were played by participants." In the system we are currently developing, intelligent agents are represented in the story world by a set of goals and potential actions that reflect their personalities. The drama manager is constituted by a set of rules directly extracted from RPG practices. These rules, implemented within the agents' personalities and goal structures, but also within the world framework, are triggered when the performance requires them. They have been designed to respond to the intensity of the performance and monitor the activity and emotional state of each character in the performance.

5 FearNot! An Example of Emergent Narrative

The project VICTEC (2002-05), involving five partners in the UK, Germany and Portugal, applied virtual dramas acted out by 3D graphically-embodied characters to anti-bullying education [7,18]. The aim of the FearNot! (Fun with Empathic Agents Reaching Novel outcomes in Teaching) demonstrator was to allow children to explore what happens in bullying in an unthreatening environment in which they took responsibility for what happened to a victim, without themselves feeling victimized. The child was asked to act as an 'invisible friend', and to give coping advice between episodes which would influence the behaviour of the victim in the next one without undermining its autonomy of action and the child's ability to believe in it as a character with an independent inner life. This interactional structure described in details in [7] (ie. Agent personalities, implementation) was inspired by the Forum Theatre approach of by Boal [9] and the child acted as a spect-actor.

Given there are around 7 different pieces of coping advice a child could give, and the order in which they are given before the second or third episode would also have to be taken into account, a branching narrative of the type used successfully in MRE [13] or Carmen's Bright IDEAS [12] seems infeasible. Thus EN, in which action is driven by the characters themselves, is a natural solution to making the victim responsive to the advice the child gives. At the same time, the repetitive nature of bullying, and the fact that it is naturally episodic, does not require too much from the emergent mechanism in terms of dramatic complexity or length. The use of the Forum Theatre approach also means that the emergent mechanism does not have to take user actions directly into account. Thus this was a small-scale application of the EN concept, in which the emotional reactions of the characters fed into their individual action-selection mechanisms. Given also the indeterminacy of physical actions in the system – a character who was pushed might or might not fall over – the EN concept produced a variety of storylines in episodes with outcomes essentially rather unpredictable.

A Stage Manager was required because a choice had to be made about where each new episode was located and which characters were involved in it, as well as any other initial conditions. In addition, there has to be some method of determining when an episode has finished once there is no script encoding this information. Deciding the termination condition is a small-scale example of the performance-monitoring activity discussed more generically above.

6 Conclusion

In this paper, we presented the formulation of an emergent narrative approach to interactive storytelling as a potential solution to the recurrent problem of the narrative paradox. We have highlighted the needs for a better consideration of stories, characters, performances and narrative displays in regard to interactivity. We have also introduced the FearNot! demonstrator, we are proposing to use for future scale up and drama-oriented experiments on the emergent narrative concept. Finally, we have presented our strong position on the role we believe the user should play in regard to IS/VS. Our view on the matter reflects our belief on the form in which interactivity should also be considered in such systems and our approach of the interactivity issue from an internal-ontological approach as described in Ryan [19].

References

[1] Aylett, R.S. (1999) “Narrative in Virtual Environments: Towards Emergent Narrative”. Papers from the 1999AAAI Fall Symposium, Technical report FS-99-01, AAAAI Press, Menlo Park, 1999 pp83-86.

[2] Aylett, R.S & Louchart,S (2003) Towards a Narrative Theory for Virtual Reality. (2003) *Virtual Reality* vol 7 no 1 Dec 2003

[3] Aristotle. 330 B.C. The Poetics of Aristotle, translation and commentary by Stephen Halliwell. Duckworth, 1987.

[4] Louchart S, Aylett R (2004) Emergent narrative, requirements and high-level architecture.SETN04 2004.

[5] Louchart, S, Aylett, R.S (2003) Solving the narrative paradox in VEs - lessons from RPGs_in: Intelligent Virtual Agents, 4th International Workshop IVA2003 eds T. Rist, R.Aylett, D.Ballin, LNAI 2792 Springer 2003 ISBN: 3-540-20003-7 pp244-248

[6] Aylett, R.S & Louchart,S (2003) Towards a Narrative Theory for Virtual Reality. (2003) Virtual Reality vol 7 no 1 Dec 2003

[7] Louchart, S; Aylett, R.S; Paiva, A and Dias J. (2005) Unscripted Narrative for affectively driven characters, First International Conference on Artificial intelligence and Interactive Digital Media 2005

[8] Cliff, D & Grand. S. (1999) "The creatures global digital ecosystem," *Artificial Life*, vol. 5, pp. 77--94, 1999

[9] Boal, Augusto. (1979) The Theatre of the Oppressed. New York: Urizen Books, 1979

[10] Michael Mateas, Andrew Stern, Structuring Content within the Facade Interactive Drama Architecture AIIDE, 2005

[11] Cavazza.M, Charles.F, Lozano.M, Mead.S, Bisquerra.A: Planning formalisms and authoring in interactive storytelling, TIDSE. 2003

[12] Marsella, S.C., Johnson, W.L., & LaBore, C. (2000). Interactive pedagogical drama. In *Proceedings of the 4th International Conference on Autonomous Agents*, 2000.

[13] Swartout, W and al. Toward the Holodeck: Integrating Graphics, Sound, Character and Story. *Proceedings of 5th International Conference on Autonomous Agents* (Montreal, Canada, June 2001).
[14] Riedl M, Young M, From Linear Story Generation to Branching Story Graphs, First International Conference on Artificial intelligence and Interactive Digital Media, 2005
[15] Aylett, R.S. (2000) Emergent Narrative, Social Immersion and "Storification" Proceedings, Narrative Interaction for Learning Environments, Edinburgh, 2000
[16] Schoppers M. Advanced Decision Systems - Universal plans of reactive robots in unpredictable environments. 1987; Proceedings of IJCAI 1987
[17] Grutbildning - http://www.grutbildning.to/
[18] Paiva, A; Dias, J; Aylett, R.S; (2005) Learning by Feeling: Evoking Empathy with Synthetic Characters. *Applied Artificial Intelligence*, 2005. **19**(3-4): p. 235-266
[19] Ryan ML; (2001) Beyond Myth and Metaphor: The case of narrative in digital media-Game studies - the international journal of game research – Volume. 2001

Motif Definition and Classification to Structure Non-linear Plots and to Control the Narrative Flow in Interactive Dramas

Knut Hartmann, Sandra Hartmann, and Matthias Feustel

Department of Simulation and Graphics ,
Otto-von-Guericke University of Magdeburg,
Universitätsplatz 2, D-39106 Magdeburg, Germany
Knut.Hartmann@isg.cs.uni-magdeburg.de,
Sandra.Hartmann@web.de, Matthias.Feustel@gmx.de

Abstract. This paper presents a visual editor which supports authors to define the narrative macrostructure of non-linear interactive dramas. This authoring tool was used to represent Propp's narrative macrostructure of Russian fairy tales in non-linear plot graphs. Moreover, Propp's thorough characterization of basic narrative constituents by explanations, discussions and examples of their different realizations in his corpus is utilized to construct an automatic classification model. A semi-automatic classification supports (i) authors to associate new scenes with basic narrative constituents and (ii) players to control the narrative flow in the story engine. For the latter task, the selection of an appropriate plot element and behavioral pattern within the dialog model in response to player interactions is controlled by similarities between stimuli and known realizations of basic narrative constituents or behavioral patterns. This approach tackles the main challenge of interactive drama — *to balance interactivity and storyness.*

Keywords: Interactive Drama, Non-linear Plots, Authoring Tools, Propp Functions, Motifs, Document Classification, WordNet, Narrative Control, ChatterBots.

1 Introduction

The research in interactive drama focuses on the development of formalisms and techniques to select or generate interesting plot structures which integrate player interactions coherently. As authors need a profound knowledge of internal data structure of these prototypical story engines there are just a few tools available which support a collaborative authoring, (re-)structuring, and evaluation of interactive dramas by non-programmers.

This work presents an authoring tool to define and (re-)structure non-linear plot structures and proposes a story engine, both based on the notion of narrative macrostructures. The paper is organized as follows: Sec. 2 introduces the concept of narrative macrostructures. Sec. 3 presents the architecture and the individual components of our narrative authoring system. Sec. 4 discusses related work and Sec. 5 summarizes the contributions of our work.

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 158–167, 2005.

2 Narrative Macrostructures

This paper employs Propp's analysis [15] of typical plot patterns and their dependencies in a specific narrative genre. However, authors are not restricted to use this specific inventory of plot patterns.

Propp's analysis of basic narrative constituents (*functions*) is based on two abstractions:

(i) a classification of the *dramatis personae* according to their *roles* and
(ii) an evaluation of actions with respect to common effects and according to their positions within the story.

Using this approach analysts are able to extract the *macrostructure* (*deep structure*) for a given story or plot (*surface structure*). To avoid the general and ambiguous term *function*, we will refer with *motifs* to basic constituents of the deep structure (as used by Pike's [12] and Dundes' [4]) and with *scenes* or *plot elements* to basic narrative entities of the surface structure.

Propp claimed that the genre of Russian fairy tales could be defined through a small inventory of motifs which are arranged within two alternative sequences. Propp's concepts influenced research to analyze narratives with highly conventional set of roles and motifs (e.g., myths, or epics) and even other media (e.g., reality TV shows). Narrative macrostructures have also been proven to be a very useful framework for interactive dramas. However, most researchers argue that the inventory of motifs as well as their order have to be adapted for different narrative genres. Moreover, plot structures in interactive dramas have to integrate player interactions, therefore additional control mechanisms are required to ensure the coherency of non-linear plots.

In order to support the analysis of narratives, which employs the identification of plot elements and their motival classification, Propp provides a list of common motival *variants* and their prototypical realization.[1] Unfortunately, the motival analysis is both subjective and time-consuming, as the segmentation of the narrative into basic plot constituents and their motival classification involves a high amount of abstraction, transformation and comparison with the prototypical examples and variants. Moreover, assimilations between realizations of motifs complicate the analysis. Finally, different symbolic representations for Propp's motifs used within the Russian, English, and German versions of his book make it hard to discuss and share the results. However, Propp's detailed explanations, discussions and examples of different realizations of motifs can be exploited in order to construct an automatic classification model.

3 An Authoring Tool for Interactive Dramas

We developed an authoring toolkit to create and structure non-linear plots and an initial prototype of a story engine which both are based on narrative macrostructures. Fig. 1 presents the architecture of our narrative authoring tool.

The visual *plot editor* supports the definition of new motifs and their dependencies. Authors define the content of individual plot elements or scenes within the *content editor*. This process involves the assignment of scenes to motifs.

[1] Propp defined 31 motifs which frequently exhibit many variants (up to 25).

A semi-automatic classification of new scenes reduces the number of candidates which have to be considered by a human interpreter and helps to prevent incoherent classification through subjective decisions from independent human experts. Thus, more motifs and motival variants can be established without spoiling the classification process. Our authoring tool supports a collaborative story development: the overall plot structure and dependencies between plot elements, abstract scene descriptions or dialogs, sounds, and animations could be provided by specialized experts.

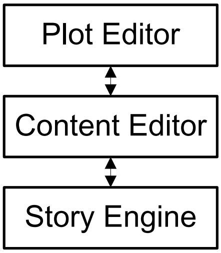


Fig. 1. Architecture

We also started to implement a *story engine*, which employs an automatic classification of player stimuli to control the narrative flow. However, this paper will mainly focus on the authoring toolkit.

The basic tasks in our narrative authoring tool comprises (i) the definition of new motifs and their arrangement within a non-linear plot, (ii) the specification of basic plot elements (scenes), (iii) the classification of scenes according to a given set of motifs, and (iv) the interpretation of player interactions within the story engine. All these tasks are described in separate sections.

3.1 Authoring Non-linear Plot Structures

The visual editor is used to define or select motifs and their causal or temporal dependencies in a *plot graph*. Our plot graph editor currently employs Propp's inventory of motifs. We defined positive and negative variants for Propp's motifs.[2] Fig. 2 presents the plot graph for an interactive drama in the fairy tale world. The branching in the subtree on the left side reflects the alternative motif sequences in Propp's formula:

$$ABC\uparrow DEFG\,\frac{HJIK\downarrow Pr-Rs^0\ L}{LMJNK\downarrow Pr-Rs}\,Q\ Ex\ TUW*$$

Note that the motifs where the story ends are visualized with thick borderlines. Authors can select the subset of motifs required for their plots and specify the dependencies between motifs interactively.

To simplify the user interface we designed a set of icons for Propp's motifs and display their textual descriptions in tool-tips. All icons follows 3 design principles:

1. a constant color-code is used to identify the roles of the dramatis personae throughout the motifs,
2. icons for positive and negative variants of motifs use the same visual elements in different colors,
3. icons for related motifs as often found in Propp's analysis reuse most of their visual elements to signal their relation.

Moreover, additional descriptions, variants, and examples of motifs establish a training set for the classification model. In the current system this information was extracted from Propp's seminal analysis.

[2] This comprises the following motifs: *beginning counteraction* C, *hero's reaction* E, *provision or receipt of a magical agent* F, *struggle* H, and *solution* N.

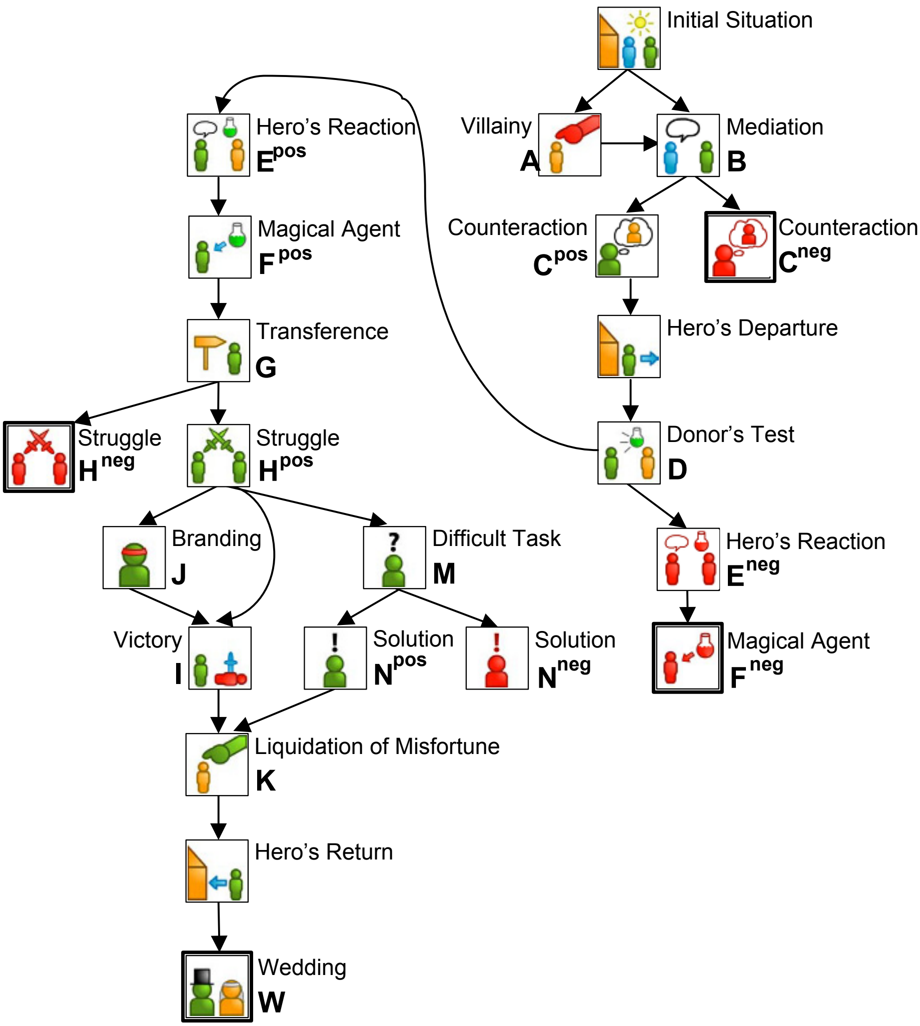


Fig. 2. A plot graph for a simple adventure game using Propp's motifs

Authors can replace Propp's motifs with another inventory of motifs (e.g., Polti's dramatic situations) or model typical patterns of social behavior, but have then to provide appropriate icons and descriptions for individual motifs and start the determination of a new classification model.

3.2 Content Editor

Scenes provide an abstract descriptions of individual plot elements. Authors can provide a textual description of the content, specify the dramatis personae and the scene location. Moreover, authors can assign behavior patterns for dialogs and interactions within a motif as well as preconditions for their successful application. This outline has to be refined by appropriate dialog sequences and interactions. Fig. 3 presents the scene editor for a plot fragment of our story.

Scenes have to be associated with motifs in order to connect the the elements of the surface and the deep structure. The proper classification of scene according to a predefined inventory of motifs is neither easy nor unambiguous as motifs can be realized in an infinite number of ways (due to abstractions, variants, and assimilations).

In order to ease this subjective and time-consuming classification, authors can exploit the classification model to select an appropriate motif. Therefore, the classifier assigns probabilities according to similarity measures between the scene and known motif descriptions. Tab. 1 lists the suggested motif classifications according to the scene description in Fig. 3.

Fig. 3. Content editor view

This semi-automatic classification supports a consistent classification of scenes and a dynamic inventory of motifs within a collaborative plot authoring tool. Moreover, the consistency of the classifications within the corpus can be evaluated.

Table 1. Suggested motival classification for a scene description

Motif	Denotation
Initial situation	α
Villainy	A
Liquidation of misfortune or lack	K
Difficult task	M
Solution (negative)	N_{neg}
Solution (positive)	N_{pos}
Reaction of the hero (negative)	E_{neg}
Reaction of the hero (positive)	E_{pos}

3.3 Motif Classification

The motif definition also comprises textual descriptions of motifs and their variants. Motif descriptions, motif variants, and plot fragments taken from an annotated corpus are considered as *instances* of a common *class*. These text fragments form a *training set* to extract parameters for a *classifier* which assigns a classification to new input data.

We employed standard techniques from *document classification* [8] in our motif classification algorithm. Textual descriptions are transformed into *document vectors*.

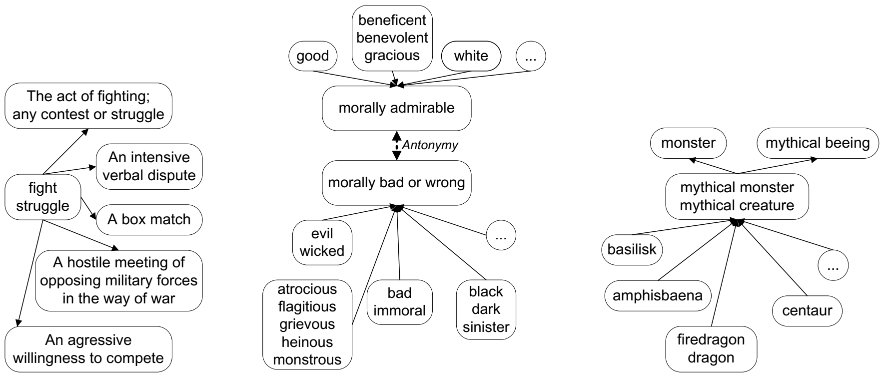


Fig. 4. Semantic fields for some actions, attributes, and dramatis personae

Therefore, the input string is segmented into tokens. Stop words are removed and morphological and inflectional variants are normalized according to Porter's stemming algorithm [14]. If the stop word list contains all determiners and preposition and inflection suffixes are removed from the remaining tokens the document vectors for the string

"The hero struggles with the evil dragon."

would contain the following set of tokens:

$$\{ \text{hero, struggle, evil, dragon} \}$$

Moreover, we employ the WordNet thesaurus to cope with semantically related lexems. WordNet [5] and its language specific variant GermaNet [6] subsume lexems that could be replaced mutually in some contexts within *synonym sets* (synsets). In other words, there is a common semantic interpretation for the lexems in a synset. Different semantic interpretations of a lexem are reflected by its contribution to different synsets. For example, the lexems *evil* and *wicked* are all member of a synset which refers to the concept of *morally bad or wrong*. This information is reflected within the middle graph of Fig. 4, which also contains other synonym sets in its nodes.

Table 2. Parts of speech, their semantic categories and lexico-semantic relations

Part of Speech	Category	Relations
Verbs	Actions, Events	Hypernymy Troponomy
Nouns	Objects	Hypernymy Meronymy Holonymy
Adjectives	Attributes	Hypernymy Antonymy

Moreover, a number of lexico-semantic relations between synsets are specified. As the individual parts of speech are associated with different semantic categories, they also obey a different inventory of lexico-semantic relations. Tab. 2 presents the parts of speech as well as their associated semantic categories and lexico-semantic relations. *Hypernymy* and *hyponymy* are the most frequent relations. They express super- and subordinations between concepts. For objects, several part-whole relations are represented: *meronymy* and *holonymy* (i.e., has-parts, is-part-of, is-member-of relation). Adjectives can express opposite attribute values (*antonymy*). Finally, *troponymy* indicate particular ways to execute actions.

Fig. 4 presents a small fragment of the graph to represent the *fighting* action (verb), the *evil* attribute (adjective) and the noun *dragon*, a typical actor for the villain role, within WordNet. The nodes contain the lexems of a synset whereas unlabeled edges represent hypernymic relations.

During a *generalization step*, the superordinate concepts for all tokens are added to the document vector. Note that this has to be done for all semantic readings as neither a part of speech tagging nor a semantic disambiguation is available in the current system. Thus, the shared superordinate concepts of semantically related lexems guarantee a minimal similarity between the document vectors for synonymic linguistic expressions.

For example, the document vectors for two plot elements of the motif *struggle H*:

The hero struggles with the evil dragon.
The hero fights with the sinister basilisk.

share the concepts *"hero"*, *"the act of fighting; any contest or struggle"*, *"morally wrong or bad"*, *"mythical monster"*, and *"mythical creature"*.

All required resources (tokenizer, stemmer, and thesauri) are available for several languages (English, German, etc.). However, we had to adopt several components to consider specific variances in the data structures between the English and German thesaurus. Moreover, the stemming algorithm also has to be applied to the entries in the thesaurus to cope with morphological variants.

We employ several classification algorithms (nearest neighbor, support vector machines) provided by the Weka data mining toolkit [21]. As the preprocessing (creation of a document vector, determination of parameters of individual classification methods) is expensive, these results are stored within a classification model of motifs. Finally, the same transformation procedure is applied for text strings before they are classified.

3.4 Story Engine: Interpretation of Player Interactions

The motival classification algorithm are also exploited in the story engine: (i) to control the narrative flow and (ii) to select an appropriate behavioral pattern for the given stimulus of the player.

The main task of the *narrative control* is to decide whether the stimuli fit best into the current motif or whether it is an indication for a context switch to a subsequent motif. Therefore, the probabilities to classify textual inputs according to the known motifs are determined automatically, whereas for other stimuli a predefined textual description is classified. To restrict the narrative flow, we consider only those motifs which can be reached from the current motif by causal or temporal links. Thus, the player could switch to all scenes reachable from the current scene if there is enough evidence to do so. This results in a flexible game play.

All characters are modeled as reflexive agents in our story engine, i.e., specific behavioral patterns (i.e., stimulus + reaction) are defined for the individual motifs. Moreover, reaction patterns can employ several media (text, sound, and animation) and specify their coordination. Behavioral patterns are represented in the AIML format [20]. The story engine incorporates a Java implementation [1] of the AIML parser.

The *pattern selection* evaluates the similarity between player stimuli within the behavioral patterns of the current narrative context. Exact matches between the input and the stimuli pattern are preferred and processed by the standard AIML interpreter. If no patterns can be applied, the similarity between the input and the stimulus is evaluated and the best pattern is selected.

3.5 Discussion

The flexibility of our authoring tool has also some implications. As authors can define new motif and delete them, scenes may have to be reclassified or to be deleted as their reference classes might not be available any more. If an author loads or stores a plot graph, the authoring tool checks whether all the scenes have been classified properly and whether all motifs do have associated scenes with them in order to guarantee the valid traversals for every path of the plot graph.

Dialogs and interactions are realized by behavior patterns for dramatis personae as well as for items. Contextual dependencies within dialogs have to be realized through setting and accessing global variables. Finally, a visual representation for the actions and locations has to be specified by defining sprites or animations as well as background graphics or view-points within a 3D scene. Note, that the dialogs provided could considerably improve the classification model of motifs.

The plot graph offers several options to adopt the plot traversal according to the player interaction: (i) by choosing different branches in the plot graph, (ii) by choosing alternative motif realizations (scenes), and (iii) through a flexible narrative flow according to the motival classification of player interactions. The consistency of the resulting story line has to be guaranteed by an appropriate arrangement of motifs in the plot graph as well as from well chosen constraints to the scene selections.

4 Related Work

The identification and extraction of narrative macrostructures and plot patterns within a genre achieved attention from various scientific communities: literature critics (e.g., Polti's dramatic situations [13]), anthropology (e.g., Levi-Strauss, Dundes), semiotics (e.g., Bremond), and psychoanalysis (e.g., Campbell's monomyth [2]) just to name a few. In text linguistics, these common structures are often described by formal grammars [3,16], document type definition (DTD) or XML schemata [7,17]. These formal grammars and structural descriptions are convincing and elegant, however, they require an expensive and subjective manual segmentation and classification.

Recently, formal inference mechanisms have been used to formalize Propp's motifs and their sequential order (e.g., using description logics [11]). These inference mechanisms could be exploited to control the game play [10].

Propp's concepts have been applied successfully to incorporate external requirements and player interactions into the plot structure of research prototypes for interactive dramas. The Theatrix project [9] employs the classification of actors with roles to characterize human and virtual agents along with their associated actions in an interactive play. In the Geist project [19], Propp's motifs are associated with locations (stages) so that the player could interact within an Augmented Reality (AR) environment to explore historic events through their movements and interactions. Moreover, an interactive plot editor for authors without computer science skills was developed in that project. However, the editor relies on a manual classification of plot elements according to the fixed inventory of Propp's motifs [18] as opposed to our approach which enables a flexible inventory.

5 Conclusion

In this paper we presented a visual authoring tool to create narrative macrostructures and to define the resources required to design an interactive drama collaboratively. The plot graph editor enables an author to define new motifs and their dependencies. This is a requirement to adopt the structural approach for genres with a greater inventory of motifs and a less strict order of motifs. Moreover, we presented a novel approach to classify motifs according to the example set (provided by Propp's famous structural analysis or a new inventory defined by dramatists or game designers). This (semi-)automatic classification supports the analysis of narratives as well as the implementation of a flexible game engine for interactive dramas.

References

1. N. Bush. Program D, Version 4.5, 2005. http://aitools.org/downloads/.
2. J. Campbell. *The Hero With a Thousand Faces*. Princeton University Press, 1948.
3. B. Colby. A Partial Grammar for Eskimo Folktales. *American Anthropologist*, 75:645–662, 1973.
4. A. Dundes. From Etic to Emic Units in the Structural Study of Folktales. *Journal of American Folklore*, 75:95–105, 1962.
5. C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.
6. C. Kunze and L. Lemnitzer. GermaNet - Representation, Visualization, Application. In *3rd Int. Conf. on Language Ressources and Evaluation (LREC 2002)*, pages 1485–1491, 2002.
7. S. Malec. Proppian Structural Analysis and XML Modeling. In *Proc. of Computers, Literature and Philology (CLiP 2001)*, 2001. http://clover.slavic.pitt.edu/~sam/propp/theory/propp.html.
8. C. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, 1999.
9. A. Paiva, I. Machado, and R. Prada. You Cannot use my Broom! I'm the Witch, you're the Prince: Collaboration in a Virtual Dramatic Game. In *Intelligent Tutoring Systems, 6th Int. Conf. (ITS 2002)*, 2002.

10. F. Peinado and P. Gervás. Transferring Game Mastering Laws to Interactive Digital Storytelling. In *2nd Int. Conf. on Technologies for Interactive Digital Storytelling and Entertainment*, pages 48–54, 2004.
11. F. Peinado, P. Gervás, and B. Díaz-Agudo. A Description Logic Ontology for Fairy Tale Generation. In *Forth Int. Conf. on Language Resources and Evaluation: Workshop on Language Resources for Linguistic Creativity*, pages 56–61, 2004.
12. K. Pike. *Language in Relation to a Unified Theory of the Structure of Human Behaviour*. Mouton, The Hague, 1967.
13. G. Polti. *Thirty-Six Dramatic Situations*. Kesslinger Publishing Company, 1916.
14. M. Porter. An Algorithm for Suffix Stripping. *Program*, 14(3):130–137, 1980.
15. V. Propp. *Morphology of the Folktale*. University of Texa Press, Austin, 1968.
16. D. Rumelhart. Notes on a Schema for Stories. In D. Bobrow and A. Colins, editors, *Representation and Understanding: Studies of Cognitive Science*, pages 211–236. Academic Press, New York, 1975.
17. G. Scali and G. Howard. XML Coding Of Dramatic Structure For Visualization. In *Museums and the Web 2004*, 2004. http://www.spacespa.it/studi_ricerche/pdf/XML_Coding.pdf.
18. O. Schneider, N. Braun, and G. Habinger. Storylining Suspense: An Authoring Environment for Structuring Non-Linear Interactive Narratives. *Journal of WSCG*, 11, 2003.
19. U. Spierling, D. Grasbon, N. Braun, and I. Iurgel. Setting the Scene: Playing Digital Director in Interactive Storytelling and Creation. *Computers and Graphics*, 26(1):31–44, Feb. 2002.
20. R. Wallace. *The Elements of AIML Style*. ALICE A.I. Foundation, 2003.
21. I. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, 1999.

INSCAPE: Storymodels for Interactive Storytelling and Edutainment Applications

Stefan Göbel, Felicitas Becker, and Axel Feix

ZGDV Darmstadt e.V., Digital Storytelling Department,
64283 Darmstadt, Germany
{Stefan.Goebel, Felicitas.Becker, Axel.Feix}@zgdv.de
http://www.zgdv.de/distel/

Abstract. Techniques and rules, that promise a particularly effective way to tell stories, have been developed over hundreds of years. Hereby, from a technical point of view, various storymodels or story patterns might be derived from these approaches to create and tell new stories.

Within the first phase of the INSCAPE[1] project, several of these models origined in different genres such as film, theatre or fairy tales have been analysed with respect to its possible usage as templates within authoring environments for interative storytelling applications.

1 Motivation

Stories have been created and told within living memory. Typical scenarios include people sitting around a fire and listing to one person telling a story about his life or people experiencing a story by reading a book, watching a film, listining to actors at theatre or guides at a sightseeing tour, playing a game or learning a language. All these examples describe the widespread use of stories and storytelling for communication and knowledge transfer in various situations and genre.

The main difference concern the distinction between purely oral narration by human beings and storytelling in the digital age using different media as well as the level of interactivity and participation of users: ‘Consuming a story’ vs. ‘interactively experiencing a story’. Of course, no digital system can provide the possibilities of a first-person narrative presented by a human being fascinating its listeners and taking into account their reactions for story continuation.

On the other hand, especially related to Interactive Digital Storytelling in the digital age it becomes obviously that many Hollywood films, thrillers, penny dreadfuls or learning courses provide similar structures or –more technically speaking- they are based on similar story patterns and storymodels.

Based on that situation, the Digital Storytelling team at ZGDV Darmstadt analysed different storymodels within the first project phase of the INSCAPE project in order to answer the following questions:

[1] INSCAPE: Interactive Storytelling for Creative People, integrated project funded by the European Commission, see acknowledgments.

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 168 – 171, 2005.

1. How can these story patterns and storymodels support an author using an authoring environment with a story editor to create a new story?
2. Which storymodels are useful for the different Interactive Storytelling domains addressed within the INSCAPE project?

The first question primarily addresses ‘non-expert authors’ since ‘professionals’ such as film producers or stage directors do not necessarily need a story template or story patterns for creating exciting and suspenseful stories. With respect to the second question, INSCAPE application domains vary from edutainment, simulation and training to animation (comics), business and marketing or life performance.

2 Analysis of Storymodels

Mellon & Webb (1997) provide the following definition for storymodels:

‘A Component, which integrates Structure, Content, Context and Development’.

Aristotle already wrote about basic principles of storytelling in the *Poetics*. His postulates are still in use today and build the basis for further models, especially in the area of scriptwriting. Aristotle’s principles (arc model with ‘exposition’, ‘rising action to climax’ and ‘denouement’) and the closely connected pyramidal model by Gustav Freytag [3] are enormously significant for drama, too.

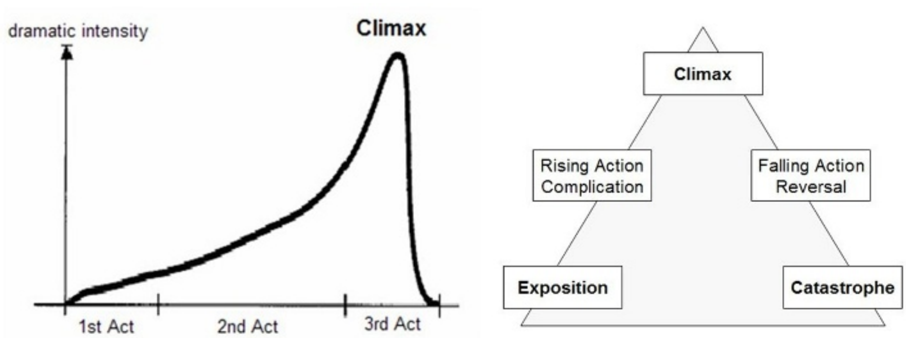


Fig. 1. Aristotele and Pyramidal Model

Syd Field [2], a script consultant in Hollywood extended the Aristotle Model with regard to its usage for film scripts. Hereby, script pages (= film minutes) are used for temporal structuring. Tobias [11] provides 20 master plots. With respect to fairy tales, the Russian formalist Vladimir Propp [8] analysed hundreds of Russian fairy tales and extracted 15 morphological functions/components appearing in all these stories. Further, Propp defined characters (*Dramatis Personae*) representing rules within the stories, e.g. an enemy, a hero, a magic agent (helper) or a princess (prize/award).

Joseph Campbell [1] analysed the myth of the hero in different kinds of stories from various regions and times. In all his stories he found the same storymodel, which he later called the hero’s journey or *monomyth*. Christopher Vogler [7] took Campbells monomyth and adopted it into a manual for scriptwriters. *'Modern heroes may not be going into caves and labyrinths to fight mythical beasts, but they do enter a Special World and an Inmost Cave by venturing into space, to the bottom of the sea, into the depths of a modern city, or into their own hearts.'*

The story of the hero's journey begins in the odinary world of the hero. The hero is introduced in his every-day surroundings. An incident, that leads to the action is depicted in all popular storymodels. Linda Seger [9] writes about a catalyst, while Robert McKee [6] quite simply calls it the *'inciting incident'*. Other terms are *'point of attack'*, *'impetus'*, or in the hero's journey, the *'call to adventure'*. In any case, the calling has to prevent the hero from living his life the way he used to. How this happened is left to the author.

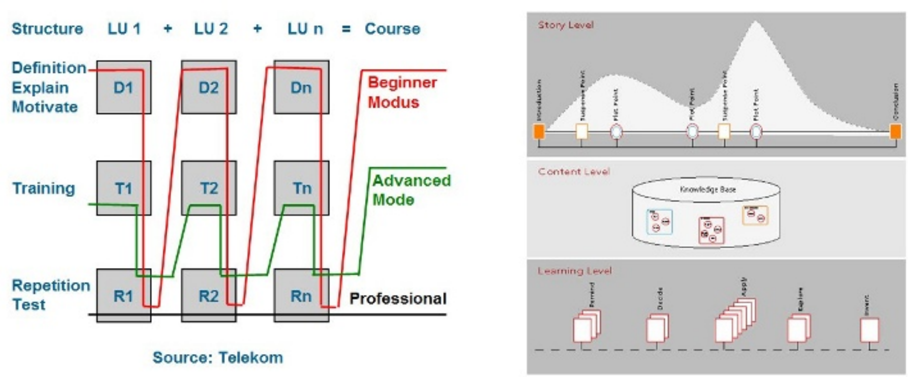


Fig. 2. Learning Courses – Linear Structures versus story-based three level concept

With respect to learning applications, the Storytelling group at ZGDV provides an integrated approach for storytelling based education purposes consisting of three levels: Story level, knowledge level and learning level [5]. The basic idea is to combine traditional learning patterns and course elements such as the definition and explaination, training or testing phases with a storymodel providing an arc of suspense and dramatic elements.

3 Storymodels in INSCAPE

Based on the defined user requirements and domain knowledge of the INSCAPE end users, available technology of the technology partners and recommendations of the research partners, the INSCAPE consortium selected the edutainment domain (Mozart Story) as first application scenario for the α-release of the INSCAPE system.

Hereby, the different storymodels described above are provided to the user in the authoring environment in the form of story templates. With respect to story characters (such as the hero of a story) a character editor provides the possibility to select predefined virtual characters (corresponding to loaded story templates, representing a specific role of the storymodel) or to define new characters (assigning characteristics, behaviour, emotions and geometry/appearance) [4]. Further on, an interaction and device editor enables the author to define user interactions which cause the transitions between the scenes of the storymodels during run-time.

As a first outcome of preliminary tests at ZGDV, the Propp model and the Hero's journey seem to build a good basis for rallye-like and *adventure-driven* mobile edutainment applications, whereas easier structures better match the needs of shorter story sequences or individual scenes (e.g. interactive installation in a museum with video recognition user interaction with a painting via pointing gestures [10]).

Another more global outcome as a result of the end user requirement analysis by the INSCAPE end users concerns the wish for a free-text story editor facility (to be realized as notes editor without any predefined structures or templates) for creating stories 'from scratch'.

A comprehensive evaluation of the INSCAPE alpha release will validate the preliminary results described above and point out requirements for further research and development in order to answer the questions how storymodels can support authors to create interactive stories and which storymodels are much appropriate for the broad spectrum of Interactive Storytelling domains.

Acknowledgments

This research is performed in the frame of the INSCAPE Integrated Project (EU RTD contract IST-2004-004150) which is funded by the European Commission under the sixth Framework Programme. More information on the project is available at www.inscapers.com.

References

1. Campbell, J.: The Hero with a Thousand Faces. Princeton University Press, Princeton, (1973)
2. Field, S.: The Screenwriter's Workbook. Dell Publishing Company, New York (1984)
3. Freytag, G.: Die Technik des Dramas. Autorenhaus Verlag, Berlin (2003)
4. Göbel, S., Schneider, O., Iurgel, I., Feix, A., Knöpfle, C. and Rettig, A. Virtual Human: Storytelling and Computer Graphics for a Virtual Human Platform. Springer LNCS (2004)
5. Hoffmann, A., Göbel, S., Schneider, O. Storytelling Based Education Applications. E-Learning and Virtual Science Centers, Idea Group Publishing, Hershey, US, 2005.
6. McKee, R. Story: Substance, Structure, Style and The Principles of Screenwriting. New York: Regan Books (1997).
7. Vogler, C.: The Writer´s Journey. Mythic Structure for storytellers and Screenwriters, Michael Wiese Productions, California (1992)
8. Propp, V.: Theory and History of Folklore. Manchester University Press, Manchester (1984)
9. Seger, L.: Creating Unforgettable Characters. Henry Holt & Company, Inc. (1990).
10. Spierling, U., Iurgel, I. (2003): „Just Talking About Art"– Creating Virtual Storytelling Experiences in Mixed Reality. Springer LNCS (2003)
11. Tobias, R.B. 20 Masterplots. Frankfurt am Main: Zweitausendeins. (1999)

Meta-Data for Interactive Storytelling

Norbert Reithinger, Elsa Pecourt, and Mina Nikolova

DFKI GmbH Stuhlsatzenhausweg 3,
D-66123 Saarbrücken,Germany
{reithinger, pecourt, nikolova}@dfki.de

Abstract. In this contribution we demonstrate how metadata for storytelling is represented using Semantic Web technologies. Three examples introduce the use of ontologies for the modelling of a database for story elements, for narratie structures, and for the domain of a game.

1 Introduction

Interactive stories are not ready-made, linear texts but rather "story spaces" representing a set of characters, together with, e.g., their personalities and motivations, as well as a set of actions and events. They are related to each other by causal and temporal relations. To provide authors the possibility to create interactive stories intuitively we need authoring tools that alleviate the process. Such a tool should provide the author with building blocks, at different granularities, to create "story spaces", and a means to ensure coherence and narrative relevance. We also need the possibility of reusing already created story elements. The main paradigms in interactive storytelling, the plot-based approach (cf. [2]), [3] and the character-based approach ([4], [1]) both require this sort of meta-data. In the EU-supported integrated project INSCAPE (`www.inscapers.com`) (FP6 IST ref. 4150), and the project VirtualHuman (`www.virtual-human.org`), funded by the German Ministry for Eduaction and Research, we use technologies currently developed in the framework of the Semantic Web activity (`www.w3.org/2001/sw`) to represent meta-data that can be used as these building blocks. In this contribution we briefly introduce technologies we use and present examples of "ontologies at work" for interactive storytelling.

2 Ontologies and the Semantic Web

The Semantic Web provides a framework that allows data to be shared and reused across application, enterprise, and community boundaries. One core technology are ontologies which are common meta-data vocabularies that explicitly represent the semantics of data. They specify the objects, concepts and entities that exist in some area, and the relations among them. They contain hierarchies of concepts and may also include sets of axioms that define the ontological concepts using a semantic formalism. An ontology which is not tied to a particular

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 172–175, 2005.

domain but attempts to describe general entities is known as a foundational ontology or upper ontology, whereas domain ontologies cover a specific application domain. Core technologies we use to represent and modify meta-data for stories are the Web Ontology Language OWL and the editing tool Protégé. OWL explicitly represents the meaning of terms in vocabularies and the relationships between those terms. It consist of a vocabulary for describing objects and properties of objects, relations between objects (e.g. disjointness), cardinality (e.g. "exactly one"), equality, typing of properties, characteristics of properties (e.g. symmetry) and many more. In addition to the representation, there exist standard tools to interpret this data automatically, e.g. to draw inferences. One of the most popular is the Jena package (`jena.sourceforge.net`). Protégé is a tool which allows users to construct domain ontologies, customize data entry forms, and enter data. It is also a platform which can easily be extended to include graphical components such as graphs and tables, media such as sound, images, and video, and other formats.

3 Using Ontologies for Interactive Storytelling

Ontology as database model: An ontology can be used to structure the content of a database in the authoring process. It provides the vocabulary to annotate the contents of the database with meta-data and models relations between the elements in the database. The DB ontology of INSCAPE (see Fig. 1 for a partial view) defines visual and audio elements that an author may want to

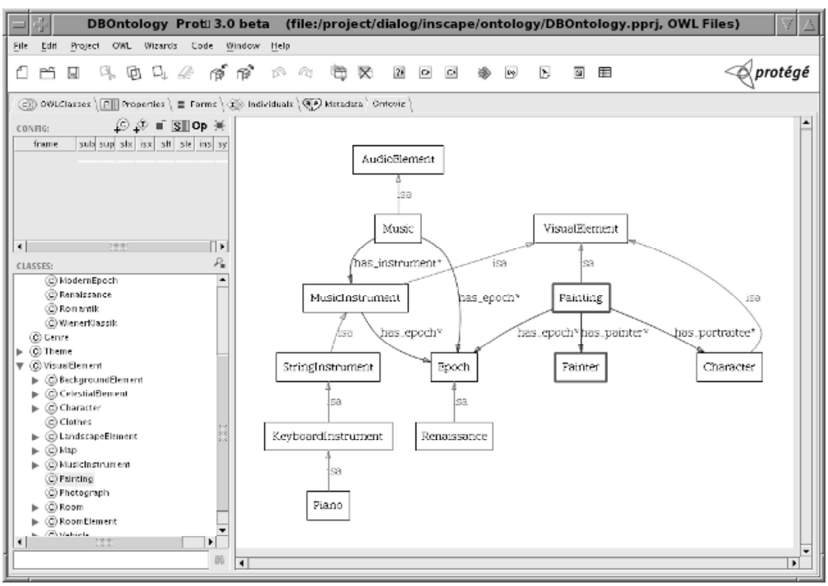


Fig. 1. Protégé display of parts of the DB ontology

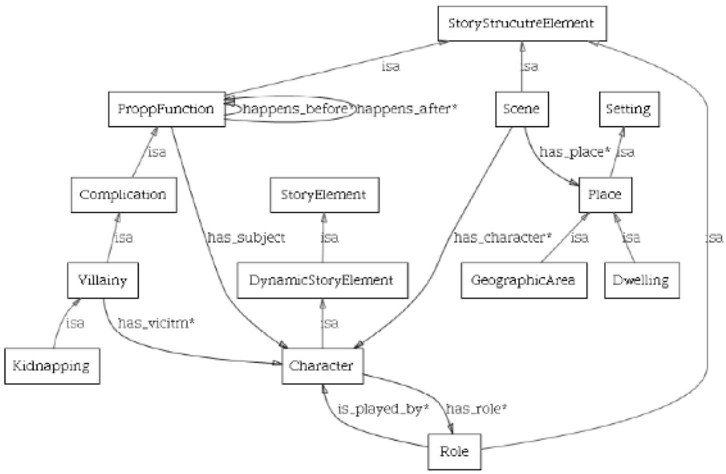


Fig. 2. Parts of the INSCAPE ontology

include in his story. They can vary according to the style chosen by the author: stills, drawings, fonts and texts, 2D animations, cartoons, movies, 3D scenes, avatars, as well as music and diverse sounds. Using the concepts defined in the ontology, story creators search the database for the object they want to participate in the story. Relationships between classes and properties of classes are used for more specific queries. For example, if an author wants to include a drawing from a particular painter, with the *hasPainter* property of the class *Painting* he gets only these entries in the database that correspond to the particular painter.

Ontology for the creation of stories: The INSCAPE ontology formalizes different narrative theories, e.g. the well-known Propp theory, which suggests that stories can be built by arranging story units according to precise rules (see also [5] for a comparable approach). The ontology thus provides the necessary knowledge source for different forms of narrative control. It facilitates the mapping between the actions of characters and users to narrative units that form the desired story structure. The Proppian functions are defined in the INSCAPE ontology in the class *ProppFunction* (see Fig. 2). The *happensBefore* and *happensAfter* properties describe the implicit dependencies between the functions. For example, the function *Mediation* cannot be applied, before the villain has caused harm or injury to some character. The ontology defines the characters to fill the roles in the story, which include *Hero*, *Villain*, *Helper*. Different locations, where the story can evolve, are also represented in the ontology, as well as the different objects that may appear in the story.

Ontology for the description of system specific domains: Clue is a Cluedo-like interactive game developed using technology from the VirtualHuman project. Four characters, three virtual and the user, try to find the murder

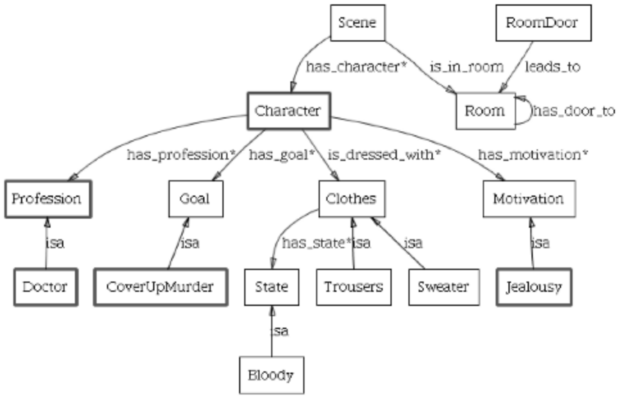


Fig. 3. Parts of the Clue Ontology

of a fifth character. The action takes place in four rooms and the user becomes to see only the room in which his character is situated. The Clue ontology describes concrete and abstract objects like persons, furniture, alibi, and motivations for murder. All actions that characters can execute (walk, search for evidences, open, close, etc.) are also formalized in the ontology. The communication between characters is defined by dialogue acts, like questions, answers, statements, accusals, and dialogue games, which are sequences of dialogue acts. The actions are defined with their roles, which are filled by ontological objects. For example, the action *Take* has as roles the object which is to be taken and the subject, who takes the object. For a more detailed description of the Clue game see [6].

References

1. Charles, F., Mead, S., and Cavazza, Ch. Character-nased Interactive Storytelling, IEEE Intelligent Systems, pp. 17-24, 2002.
2. Machado, I., Paiva, A., and Brna, P. Real Characters in Virtual Stories. In Proc. of the international Conference on Virtual Storytelling. Springer, London, pp. 127-134, 2001.
3. Grasbon, D., and Braun,N. A Morphological Approach to Interactive Storytelling. In Proc. CAst01, Living in Mixed Realities, Bonn, 2001.
4. Bickmore, T. and Cassell, J. Small Talk and Conversational Storytelling in Embodied Interface Agent. In Proc. AAAI Fall Symposium, "Narrative Intelligence", Cape Cod, MA., 1999.
5. Peinado, F., Gervás, P., and Díaz-Agudo, B. A Description Logic Ontology for Fairy-tale Generation. In Procs. of the Workshop on Language Resources for Linguistic Creativity, LREC'04, pp. 56-61, Lisbon. 2004.
6. Löckelt, M. , Pecourt, E., Pfleger, N. Balancing Narrative Control and Autonomy for Virtual Characters in a Game Scenario. In Proc. of the INTETAIN 2005, Madonna di Campiglio, 2005.

New Ways of Narrative

Embodied Reporting Agents as an Approach to Creating Narratives from Live Virtual Worlds

Ella Tallyn, Boriana Koleva, Brian Logan, Dan Fielding,
Steve Benford, Giulia Gelmini, and Neil Madden

Mixed Reality Lab, School of Computer Science, University of Nottingham

Abstract. The most common approach to creating interactive narrative involves interactive experiences which take place within the constraints of a previously constructed story. In this paper we explore an alternative approach in which participants in a virtual world, e.g., a game, simulation or large online community improvise events. These events form the raw material for the subsequent creation of narrative sequences. Building on the theoretical concept of *narrative voices* — fictional personas that deliver information in narrative form — we suggest some new approaches to creating narratives from live events. We then present one such approach, embodied reporting agents, in which automated non-player characters inhabiting a virtual world report on ongoing events to editor agents. The editor agents, in turn, compile their information and pass it to presenter agents who ultimately narrate it to external viewers. We sketch how such 'witness-narrators' can be used to investigate creation of tension and drama in the interactive story world.

1 Introduction

The main problem facing the creation of meaningful and engaging narratives that are also interactive is that interactivity disrupts the effective use of traditional narrative devices; the more interactivity for the participant, the less control over the narrative techniques and meaning the author has. Techniques to try and solve this have included methods of limiting interactivity and guiding participant actions. This can be done, for example, by only giving the participant the illusion of controlling the narrative [13], trying to overtly persuade or guide players to follow the designed action [8, 15], making the experience a game and setting the player specific goals to achieve [24], and having developers working in real time to try and create interactivity for the players depending on what they do, e.g., MMOG's like Eve [2]. Most successful interactive narrative experiences such as those found in commercial games tend to employ a combination of these methods.

In general, these methods adopt a common approach where a set of story constraints is created first, and then interactive experiences take place within these. This tends to require the creation of multiple story threads and possibilities, where the number of possibilities created is directly related to the level of interactivity experienced by the participant. Generative and emergent approaches to creating narrative have also been explored. These have tended to be based upon the creation of sets of

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 179 – 188, 2005.

algorithms to generate stories and character behaviour [4, 18, 21]. However, narratives produced by machine alone tend to seem irrelevant and uninteresting due to lack of context and style, and have not made use of or enabled high levels of audience interactivity [7]. In this paper we explore an alternative approach in which real participants improvise events in a virtual world (e.g. game, simulation or large online community), which provides the raw material for subsequently creating narrative sequences. In other words, we are interested in how we can create narratives from interactive virtual worlds, rather than how can we create interactive virtual worlds from narratives.

We envisage a range of potential applications for this approach including in:

- Simulation and training – narrating events from real-time simulations (e.g., emergency rehearsals, maintenance, social skills training and other educational scenarios) so as to provide feedback for participants after the event as part of debriefing, or possibly even during the event so as to support coordination.
- Television – more engaging ways of presenting computer games on television, building on a spate of recent TV game shows in which teams of contestants compete through computer games (e.g., Time Commanders [1] in the UK)
- Massively Multiplayer Online Role Playing Games (MMPORG) and online communities – creating news feeds for large online communities in order to build community, motivate players, drive events and seed future activities.

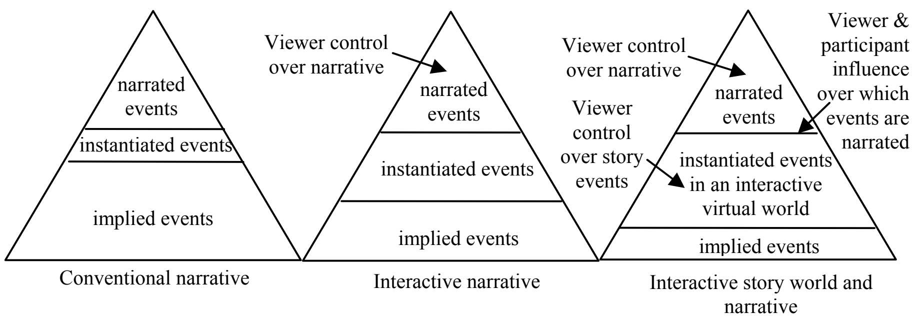


Fig. 1. Traditional and interactive narrative in relation to instantiated and implied story world

Fig. 1 compares our approach with existing traditional and interactive narratives. Conventional linear narrative (Fig. 1, left) involves the narration of story events to viewers. Many other story events will be implied but not directly narrated, e.g., the history of preceding events and the thoughts, beliefs and motivations of the characters involved [17]. A few events may actually be instantiated but not narrated, e.g., scenes may be filmed and chapters and sections may be written but subsequently edited out of the final narrated story (although these do sometimes make their way into the "director's cut" or unabridged versions). In interactive narrative (Fig. 1, middle), the viewer is able to control the narration of events, at least to some extent. E.g., they may be able to choose between different branches of a plot or to follow different characters. In turn, this may require that significantly more story events are instantiated than are narrated, e.g. entire sequences may be filmed or written that are never visited by the viewer. Here, the audience experiences a line of discourse through a set of stored possibilities [20]. In our approach that we might call 'interactive story world and narrative' (Fig. 1, right), participants interact with a virtual world in order to instantiate story events. A selection of these events is then narrated to external viewers who are not initially involved in directly interacting in the virtual world. These external viewers may then also interact with the narrative and both participants and viewers may also interact with the mechanism that determines which instantiated events get narrated and how this happens. However, even with this approach, there are still story events that are not instantiated, for example the backstory (history) behind the virtual world or the innermost thoughts and beliefs of the participants.

The remainder of this paper explores this third approach in more detail. Section 2 briefly reviews previous approaches to creating narrative from live events in virtual worlds. Section 3 then considers how narrative theory distinguishes between different narrative voices and how this suggests some new approaches to creating narratives from live events. Finally, section 4 presents one such approach – embodied reporting agents.

2 Producing Narrative from Live Events in Virtual Worlds

A number of techniques have been explored that allow action within virtual environments to be captured and edited so that a report may be included in other virtual environments or be transferred into other media.

One approach is to use virtual cameras to capture activity and broadcast in real time an edited narrative to viewers. For example, in the inhabited TV show "Out of This World" [5], there were four virtual camera operators in the virtual world who captured the action from various perspectives. Their output was fed into a conventional TV mixing desk where a professional TV director produced a video mix, which was then projected in a theater in front of a live audience. Similarly, in the improvised online drama "Avatar Farm" [9], camera operators captured different views of the unfolding drama on the virtual set, and a camera director was responsible for selecting which camera view would be web-cast live at any moment in time.

Another approach is to record the activity of an avatar. In the CAVERNsoft system [19], recording of an avatar's movements and audio is possible as part of general support for persistent virtual environments. This facility has been used to create the Vmail system [16], a form of VR email. Some multi-player 3D games also employ record and replay techniques to show highlights of previous game-play. Examples of these include FIFA soccer from Electronic Arts and the automobile game Driver from GT Interactive Software, the latter allowing players to edit together their own movies from recordings of their own actions.

Alternatively, all activity within a virtual environment can be completely recorded. For example, as part of the COVEN project, the DIVE system was extended with event logging facilities that could completely record an entire virtual environment and the activity within it. Although initially implemented to support the statistical analysis of patterns of user activity in relation to network traffic [11], this recording facility was subsequently extended to allow a previous session to be recreated (although not within another live virtual environment). The dVS system supports a similar facility for recording and then replaying a virtual environment.

The technique of temporal links [14] extends this general approach by providing a flexible mechanism for dynamically inserting past recordings of multi-user sessions within other live virtual environments. All activities within a collaborative virtual environment are recorded, including the environment itself, multiple users' movements, speech, and interactions with virtual objects. The recorded events are then recreated within a live virtual environment so participants can move around the recorded material, viewing it from any perspective. By manipulating the properties of the "temporal link" application, developers and end-users can manipulate the temporal, spatial and presentational relationships between the live environment and the recording. For example, recordings can be fast-forwarded or run backwards; the recorded material may appear to be directly overlaid on, or may be scaled relative to the live material; it may also be presented indistinguishable from the live material or it may be rendered translucent to make it ghostlike.

Various applications of the temporal links technique have been demonstrated [13], including to allow actors to enter an environment and quickly enact and record complex scenes that are then played out many times, to create flashbacks within a story, to support the post-analysis and discussion of events within virtual environments and to create offline rendered, high quality animations.

3 Narrative Voices

The techniques described in the previous section focus on generating accurate information from an interactive experience and presenting it as a review of what actually happened. Their aim is to inform an audience of interesting activities that have occurred, and as such they are conceptually similar to the presentation of news or sports in traditional media. However, in narrative fiction, there is usually a sense of a narrator's personality behind the information that is presented, a certain *someone* whose presentation and description of events is done in such a way as to produce an effect. This fictional persona who delivers the information, and the effect they may have on the information they present, is an integral part of a narrative experience, and an important authorial device.

The theory that describes the concept of different information sources or personas within a story often refers to them as 'narrative voices'. This theory is most richly described in literary fiction, and we have drawn on the theories of Rimmon-Kenan [22] and Chatman [6] to inform the concept of narrative voices within our work.

Narration can be done by many different types of fictional persona, which can range from the implied voice of the author to a character within the story. Also, it may be done by different narrators at different times in a single story. An examination of the characteristics of different narrators described in these theories shows a range of possible types of narrative voice and also reveals that some are underused in current work in this field.

Narrators may be embodied i.e. have a physical presence, or they may be omniscient. Omniscient narrators have no physical presence and can be in all places at all times. This sort of narrator is often the one whose voice we accept as the objective truth in the matters they describes, since they are generally more objective and, therefore, accurate. Embodied narrators have personalities that colour their view of what they present. For example, they often have a vested interest in showing the information in a biased way, and this may or may not be apparent to the audience. This is a popular narrative device; e.g., where a narrator may at first appear to be sympathetic to a protagonist but in the end turns out to be plotting his downfall. There are various factors that affect the particular type of information an embodied narrator provides and the reliability of his narrative, including:

- Physical – the spaces and events that the he physically has access to;
- Perceptual – what the he is able to perceive;
- Emotional – his motives and the relationship he has with events and characters;
- Ideological – the world view and value system that influences his judgment of events;
- Temporal – the passage of time between events and his narration of them, and also the influence of other events in between.

In fiction the audience is made aware of unreliability by signs or indicators in the narrative. Rimmon-Kenan describes the main sources of unreliability as: limited knowledge, personal involvement, and problematic value-schemes. There are a variety of ways in which a character-narrator's knowledge may be limited. For example, his character may be very young or too inexperienced to understand the events that are happening, or perhaps we find that he did not have sufficient access to events to see them clearly and is instead relying on his interpretation rather than certainty. A character's involvement with the plot or other characters may lead to biased information. For example, if the narrator has a particular dislike of a character his interpretation of the character's actions may be distorted by his own feelings. A problematic value-scheme generally only becomes apparent if it seems to disagree with that of the overall voice of the narrative. For example, facts may unfold in a story and prove a narrator to be wrong, many other characters' views may be different to those of the narrator, or there may be internal inconsistencies within his narration itself.

Embodied narrators can be positioned either inside or outside of the story world. If they are outside they tend to take a story-teller role. Embodied narrators that are inside the story world are called character-narrators. These fall into two distinct types, the witness-narrator and the protagonist-narrator. The protagonist-narrator is a central character who describes his experiences in the story world. This also has parallels in the work described in section 2, e.g., a central player or character that relates his experiences directly. However, the concept of the witness-narrator has yet to be explored in interactive narrative. The witness-narrator views events from within the action, but is not directly involved in furthering the unfolding plot. This means they often provide a more objective and broader view to that of the protagonist, and differences in accounts may occur according to their personality and role in the story.

4 Reporting Agents

In this section we briefly outline a new approach, *embodied reporting agents*, which offers the possibility of new forms of interactive narrative that more closely resemble the dramatic experience of a fictional narrative.

4.1 Concept

The reporting agents technique is based on the principles that agents that capture information about a 3D virtual environment are directly embodied within this environment (and so are visible and subject to the same constraints as other participants in the environment); and responsibility for extracting, filtering and reporting of information is distri buted between different types of agent [10]. Specifically, a 3D environment can be inhabited by participants (the members' avatars, capable of directly influencing the environment) and 3 types of agents – reporters, editors and presenters. Reporters are embodied in the virtual environment, but unlike the participants cannot influence it directly. Events the reporters judge significant are reported to editors. Editors have two main responsibilities. First, they pass interesting segments of the output generated by the reporters to the presenter(s). Since reporters are not infallible, this data should be verified as necessary before being passed on, e.g., by clarifying conflicting reports or by requiring multiple reporters to detect the same event. Second, editors attempt to maximise the collection of relevant and interesting information by assigning reporters in such a way as to provide good coverage of the events in the environment, e.g. by directing idle reporters to under-reported or interesting regions within the environment. Presenters are responsible for delivering information to the viewers at an appropriate time and in an appropriate format (e.g. SMS messages vs animated talking heads). Different kinds of presenters may have differing temporal relationships to the reported events, e.g. real-time vs retrospective commentary, and/or report differing degrees of detail, e.g. focusing on major events, or a more detailed summary of all events, or only on events which touch a particular character.

4.2 Implementation

We have implemented a prototype reporting agents system based on ‘Capture the Flag’, one of the game types provided by Unreal Tournament (Fig. 2). The prototype consists of a variable number of embodied reporter agents, a single (non-embodied) editor agent and one or more (non-embodied) presenter agents. We use the Gamebots interface [3] to allow agents to communicate with the UT game server. The agents themselves are implemented using the SIM_AGENT toolkit [23]. Gamebots provides each reporter with data that approximates to that available to a player. A reporter’s sensory range is limited, and to obtain information about events in other parts of the game world, the reporters must physically move to a different location. By remembering the objects in the game world that they have sensed in the past and the state of objects that they can currently sense, the reporters can attempt to infer which events are taking place within the game, and their significance. The current implementation also includes three presenter agents: an “in game” presenter that provides brief topical reports of events in the game in real time, an “IRC bot” presenter that produces periodic summaries of events in the game on an IRC channel while the game is in progress (see fig. 3), and a “post game” presenter that generates a text summary of events in the game when the game is over (see fig. 4). Viewers can direct the activities of the reporters and editors (in a general way) by interacting with the “in game” and “IRC bot” presenters. E.g., a viewer can indicate an interest in particular kinds of events (which may relate to a particular team or player). This information is used by the editors to direct reporters to events that viewers are likely to find interesting, and by reporters in deciding, e.g., which players to follow and which events to report.

Fig. 2. An embodied reporter and a player within Unreal Tournament

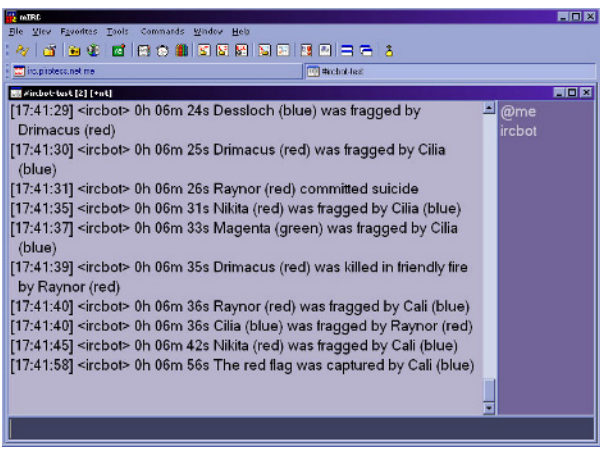


Fig. 3. Example periodic summaries on an IRC channel by an “IRC bot” presenter

The match was a Capture the Flag game on ctf-simple, lasting for 3 minutes and without any score limit. A frantic start to the match saw the red team take an early advantage, claiming the first flag capture of the game. With this seemingly spurring them on, they went on to establish a sizeable cushion between themselves and their opponents. By the end of the match, the red team had a large score advantage over their adversaries, the game ending up as a whitewash with the final scores at 5-0.
1st minute: an unknown player took the blue flag, and this resulted in a successful flag capture. The flag changed hands en route…

Fig. 4. Example summary of events by a "post game" presenter

5 Discussion

The key differences between the reporting agents approach and the techniques for producing narrative from live events in virtual worlds described in section 2 can be characterized in terms of three issues: omniscience, bias and embodiment.

- Omniscience refers to whether there is some component of the system that is able to capture a canonical view of events. i.e., a particular version that all involved can agree captures all of the essential events from a privileged central perspective.
- Bias refers to the potential of the system to deliberately bias the retelling of stories from captured events. Bias might be introduced at the recording stage or in the subsequent editing and presentation of events.
- Embodiment refers to whether and how the recording of events is visible to the participants. Are they aware that they are being recorded and are they able to react accordingly, e.g. by constraining their behaviour.

The techniques described in section 2 tend towards being omniscient, relatively unbiased and disembodied whereas reporting agents are not omniscient, may be biased, and are embodied. The concept of an embodied narrator who also has a detectable presence in the story world, physical limitations and a personality, i.e. a narrator who is a character, has been rather less explored, particularly in the role of a character-witness. By combining the three roles of reporter, editor and presenter in a single agent, we begin to approach the notion of a "witness-narrator" who is embodied inside the story world, but not in a role that is central to the plot. The characteristics of the embodied character-narrator underpin a rich array of possibilities for developing narration that is more complex than a straightforward rendition of an experience. By exploring qualities described in the theory, we can develop new possibilities for the acquisition and presentation of information from interactive events to provide not just reviews but stories.

6 Future Work

In our future work, we propose to explore how reporting agents can be employed in order to shift some of the events that are 'instantiated' by the viewer/participant into the 'narrated events' level of our diagram. In particular, we intend to investigate the role of witness-narrators in creating tension and drama in the interactive story world.

Because witness-narrators view events from *within* the action, the narrative they construct out of these events is influenced by their personalities and the context in which they view the events. When these narratives are fed back into the story-world to the viewers and participants, they inevitably influence the participants' perception of the events. This biased perception of the story-world will, in turn, have an impact on the participants' choices as to what events are to be instantiated.

This recursive process, whereby instantiated events are fed back into the story world via internal narrators' points of view, which, in turn, influence the unfolding of the story events, will be explored within a game scenario in which interactive story worlds are actively explored by the participant players. The players' choices will be influenced by the narratives produced by agents. In order to successfully proceed in the game story world, the player will need to actively seek to unmask conflicting narratives, biased interpretations, unreliable reports and elaborate their own personal view of the events, i.e. their own narrative.

Acknowledgements

This research is performed in the frame of the INSCAPE Integrated Project (EU RTD contract IST-2004-004150) which is funded by the European Commission under the sixth Framework Programme. More information on the project is available at www.inscapers.com

References

[1] http://www.totalwar.com/time2.htm - Time Commanders, verified 2nd June 2005
[2] http://www.eve-online.com/ - Eve (1997-2005) CCP, verified 2nd June 2005
[3] Adobbati, R., Marshall, A. N., Scholer, A. et al (2001) Gamebots: A 3D virtual world test-bed for multi-agent research, Proc. of the 2nd International Workshop on Infrastructure for Agents, MAS, and Scalable MAS, Montreal, Canada, 2001
[4] Bates, J., (1992) The Nature of Characters in Interactive Worlds and the Oz Project, Technical Report CMU-CS-92-200
[5] Benford, S. D., Greenhalgh, C. M., Craven, M. P., Walker, G., Regan, T., Wyver, J. and Bowers, J. (1999) Broadcasting On-line Social Interaction as Inhabited Television, Proc ECSCW'99, Copenhagen, Denmark, Kluwer, pp 179-198.
[6] Chatman S. (1978) Story and Discourse, Narrative Structure in Fiction and Film, Cornell University Press
[7] Crawford, C., (2003) Assumptions underlying the Erasmatron storytelling system. In Narrative Intelligence, Mateas, M., and Sengers. P., (eds) 189-197
[8] Doyle, P., Hayes-Roth, B., (1996) An Intelligent Guide for Virtual Environments, Knowledge Systems Laboratory Report No KSL 96-20
[9] Drozd, A., Bowers, J., Benford, S., Greenhalgh, C. and Fraser, M. (2001) Collaboratively Improvising Magic: An Approach to Managing Participation in an On-Line Drama, in Proc. ECSCW'01, Bonn, Germany, Kluwer , pp. 159-178.
[10] Fielding, D., Fraser, M, Logan, B. and Benford, S. (2004) Extending Game Participation with Embodied Reporting Agents, Proc. ACE'04, ACM Press

[11] Frécon, E., Greenhalgh, C. and Stenius, M. (1999) The DIVE-BONE – an application-level network architecture for Internet-based CVEs, Proc. ACM Virtual Reality Software and Technonology (VRST'99), London, UK, pp. 58-85.

[12] Galyean, T. A., (1995) Narrative Guidance of Interactivity, PhD thesis MIT

[13] Greenhalgh, C., Flintham, M., Purbrick, P., Benford, S. (2002) Applications of Temporal Links: Recording and Replaying Virtual Environments, Proceedings of IEEE Virtual Reality (VR 2002), Orlando, Florida

[14] Greenhalgh, C., Purbrick, J., Benford, S., Craven, M., Drozd, A. and Taylor, I. (2000) Temporal links: recording and replaying virtual environments, Proceedings of ACM Multimedia (MM 2000), L.A

[15] Holmquist, L.E., (1997) Loose Scripts and the Influence Engine: A Model for Controlling Multi-Actor Interactive, Proceedings of Writing and Computers 10, Brighton, UK

[16] Imai, T., Johnson, A., and DeFanti, T. (1999) The virtual mail system, Proc. IEEE Virtual Reality (VR'99), Houston, USA, p. 78

[17] Iser, W., (1980) Act of Reading: A Theory of Aesthetic Response, John Hopkins University Press.

[18] Lebowitz, M., (1984) Creating Characters in a Story-telling Universe, Poetics, Volume 13, No.3 pp. 171-194

[19] Leigh, J., Johnson, A., DeFanti, T., Brown, M., et al. (1999) A review of tele-immersive applications in the CAVE research network, Proc. IEEE Virtual Reality (VR'99), Houston, USA, pp. 180-187.

[20] Liestøl, G., (1994) Wittgenstein, Genette, And the Reader's Narrative in Hypertext, in Hyper/Text/Theory, (ed.) Landow, G. P., The John Hopkins University Press

[21] Perlin, K., Goldberg, A., (1996) Improv: A system for Scripting Interactive Actors in Virtual Worlds, Computer Graphics Proceeding 1996 pp. 205-217

[22] Rimmon-Kenan S. (1983) Narrative Fiction: Contemporary Poetics, Methuen & Co. Ltd

[23] Sloman A.and Poli R. (1995) SIM AGENT: A toolkit for exploring agent designs. In M. Wooldridge, J. Mueller, and M. Tambe, editors, Intelligent Agents II: Agent Theories Architectures and Languages (ATAL-95), Springer–Verlag, pp. 392–407

[24] Tallyn, E., (2000) Applying Narrative Theory to the Design of Digital Interactive Narrative with the Aim of Understanding and Enhancing Audience Participation, PhD thesis, UWE, Bristol

Telling Stories Knowing Nothing: Tackling the Lack of Common Sense Knowledge in Story Generation Systems

Hans-Georg Struck

Independent Screenwriter
georgstruck@foni.net

Abstract. To create human-level stories, story generation systems need extensive common sense knowledge about the world and human psychology. We are exploring a two-step approach to compensate for this lack of world knowledge in a joint research project at ZGDV Darmstadt Digital Storytelling Lab: First, we employ a character-driven drama model to annotate a collection of film scenes, creating an extensible, 'first-person-perspective' story grammar to substitute the system's lacking theory of mind. Second, we imbue the objects in a game world with knowledge how to achieve a dramatic effect on a protagonist within their specific environment. Combining these two sources of knowledge, we hope to create a system that is capable of generating individual educational game stories for role-playing characters in a large variety of game worlds.

1 Introduction

Most game and learning applications have been interactive storytelling systems. Based on pre-existing stories, interactive storytelling systems organize story-playback in a non-linear fashion. In contrast, story generation systems are faced with the task of 'inventing' a new story autonomously. Only very few story generation systems have been built to date, and for a reason. In order to truly create an interesting story, a story generation system would need to know everything a grown-up human knows.

In a simplified description of a writer's imaginative thought process, a writer sets a character up against a dramatically interesting situation, asks himself how the character would react, and then goes back to the situation to see how the world could in turn react to keep the situation interesting. This constant back and forth between world and character, action and reaction is characterized by a complex evaluation of possibilities, usually requiring a vast amount of knowledge about human psychology and many factual domains of the world.

This issue of world knowledge has stifled the development of story generation systems since their inception. Roger Schank abandoned two decades of research on artificially intelligent story understanding because he deemed the common sense problem an insurmountable obstacle [15]. Early systems have thus confined themselves to the narrow domains of Aesopian fables (Talespin by Meehan [11]), Arthurian courtship stories (Minstrel by Turner [22]) or the common elements of Russian folk fairytales (Propp Engine by Schneider et al. [18]). The most impressive story generation system to date

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 189 – 198, 2005.

takes the principle of narrow domains to the extreme, by limiting itself to stories about betrayal in a specific college setting (Brutus by Bringsjord and Ferrucci [3]).

We are attempting to build a story generation system for educational role-playing games at the ZGDV Digital Storytelling Lab in Darmstadt and have developed two approaches to work around the limits of artificially intelligent systems in our conceptual phase. For the purposes of this article, I will address the lack of knowledge about human psychology, and the lack of knowledge about the world, although highly interrelated, separately as the theory of mind bottleneck and the common sense bottleneck. To address the theory of mind bottleneck, we employ principles of story grammar-based systems, while we address the common sense bottleneck with principles of world-based systems. [2]

In a brief overview, our story generation system shapes user input into a dramatic story in five steps: The system asks the user to define a role-playing character dramatically, suggesting example scenes with learning and development themes to the learner. Using a story grammar of themes and emotional motives from annotated film scenes, the system infers and suggests additional appropriate scenes. The system then uses a character-driven drama model to map these scenes in a three-act plot structure. Finally, the system adapts these scenes to a new setting by using ontological information from the available game world and its objects. To create a game, the chosen scene annotations are passed on to the game characters and objects, offering high-level directions for their possible dramatic functions in a scene and their emotional effects on the main character.

2 The Theory of Mind Gap

The application is based on a character-driven drama model developed by American screenwriting teacher Frank Daniel in the analysis of contemporary feature film screenplays and theatre plays [6]. Surprisingly few drama models developed for screenwriting have yet been integrated in story generation systems [10] [7] since Laura Brendal introduced dramatic story models for story generation [9].

Frank Daniel's character driven drama model has the advantage that it defines a scene emotionally through the eyes of a character. We believe that a large collection of appropriately annotated film scenes can offer a story generation system a sufficient substitute for its lacking theory of mind.

The drama model uses a simple set of descriptors that can be applied to dramatic plot on the levels of the entire story length, as well as dramatic arcs spanning separate acts, sequences or scenes. In any dramatic story (act, sequence or scene), a main character usually wants to reach a goal, struggles to overcome an obstacle on his goal-pursuit, and risks a price for failure (stakes). If the audience identifies with the character, it shares his hopes of reaching the goal and suffers his fears of having to pay the (usually painful) price, creating a tension that is played on in the middle of the story and resolved at the end of the story.

Since the drama model defines all events in a scene in relation to the protagonist's goal pursuit, describing the protagonist's strategy, the emotional effects of a scene's outcome, and possible strategy changes, this strictly goal-related appraisal of a scene shows strong similarity to Schank's model of narrative intelligence organizing human

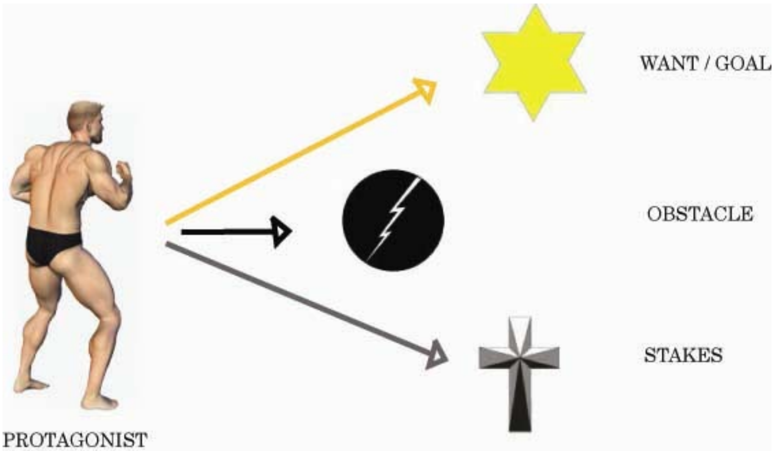


Fig. 1. Frank Daniel's character-driven drama model

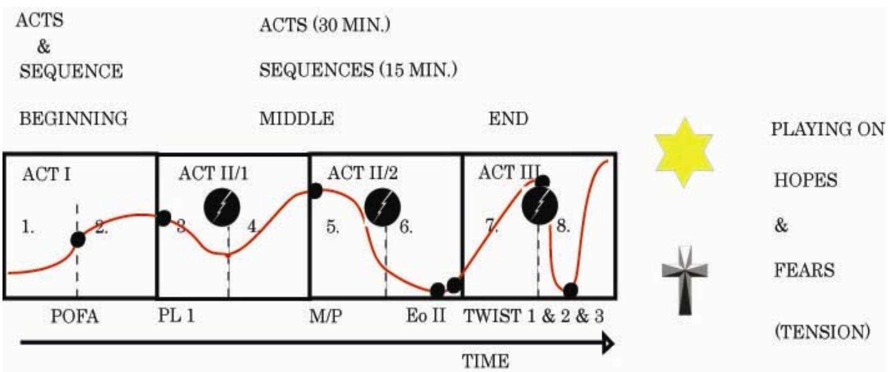


Fig. 2. The character-driven drama model mapped to a 3-act structure

memory and learning [14] [16] [20]. It will be an interesting topic for further research whether the scene database can also be used as a simplified 'narrative' memory of a virtual human character by adding the ability to talk about these 'experiences' from a first person perspective.

Frank Daniel's description of screenplays and theatre plays has been further developed by Linda Seger in a dramaturgy of character-driven screenplays [19]. A similar approach to character-centred drama has been developed by Andre Horton [5]. In this view, the character meets figurations of his worst fears and traumatic scenes in the outside world, as well as figurations of his goals in life. Every encounter actually takes him deeper towards his own hopes and fears inside. When character-driven drama models are used for story development, plot is usually generated from 'within' the character. Based on the dramatic principle that the worst fears and highest hopes are met towards the end of the story, the dramatic definition of the character is

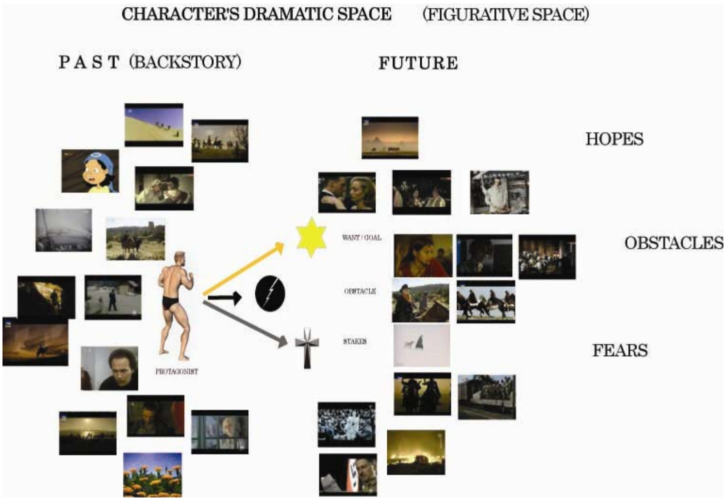


Fig. 3. An emotional definition of the character with scenes playing on his hopes and fears

mapped to a sequential unfolding in time. Plot is generated by building backwards from highly emotional key scenes in the third act, and preparing them with anticipatory scenes and similar scenes of less dramatic intensity.

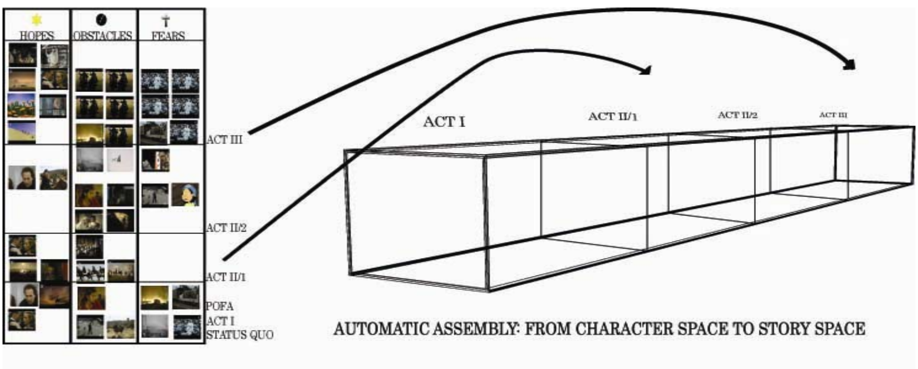


Fig. 4. Scenes evoking the highest ranking hopes or fears are mapped towards the end

Daniel's model of dramatic plot also offers a description of the learning process inherent in stories. It is a dramaturgical model that is especially strong for stories with a distinct character arc, built around a strong character transformation. Described as the protagonist's progression from an initial goal (want) to a new, more successful goal or strategy (need), the model can formally represent the learning of a protagonist in a story. Typically, this learning is concerned with soft skills.

The chosen model of character transformation can also be employed to describe the learning of hard skills (facts, procedures). Daniel's distinction of a want-goal path and a need-goal path in effective, character-centred screenplays bears a striking, fruitful resemblance to Roger Schank's description of an expert path and a failure path in his model of a successful learning episode [17]. The obstacle of a scene can thus force the learner to choose an appropriate soft skill strategy, as well as any hard skill or fact-based strategy to overcome the obstacle and reach his goal. No other dramatic model offers this parallel to the educational paradigm of learning by doing and problem-based learning. We hope to show that Daniel's narrative model is therefore well suited to model and generate stories for learning purposes.

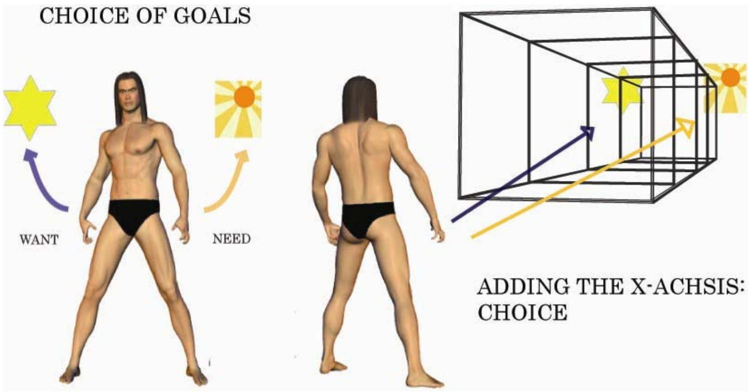


Fig. 5. The character's learning process: switching from want-goal to need-goal

Users initiate the story generation process by defining a character's hopes, obstacles and fears using example scenes. They can also users can set preferences for the learning skills and expertise they want their characters to master, deciding from example obstacles what would be the hardest (most-feared) and the easiest (hoped-for) tasks for their characters. The system maps these then presents the user with some examples how her character design would translate into an Aristotelian (dramatic, finite, catharsis-based) story. A large variety of plots can be generated from a single character design.

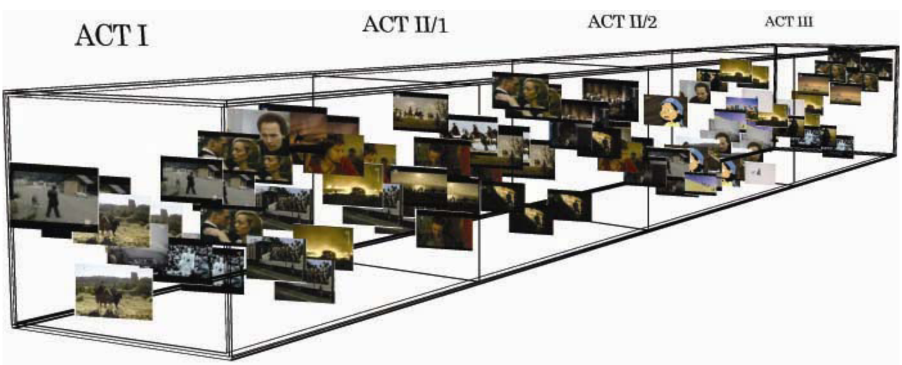


Fig. 6. Visualizing the impact of scenes on the character's emotions and learning process

3 The Common Sense Bottleneck

World-based models of story generation have tackled the common sense bottleneck most effectively so far and have been employed as the basis for all recent systems.

Instead of providing the story generation system with exhaustive knowledge about the real world, a world-based story generation system relies on a simplified, simulated model of the world composed of autonomous, goal-directed actors (agents). [11] [12] Some recent story generation systems rely upon the autonomous interaction of goal-based character agents to create an emergent narrative [4] [1], while other researchers have built 'director' or 'drama' agents that direct the autonomous actors to cooperate in goals of dramatic effect [21] [23] [10] [13].

Our application expands the world-based approach by imbuing all objects in the game with knowledge how to create a dramatically and educationally effective scene for a specific player in their environment. Anja Hoffmann from the ZGDV Digital Storytelling Lab has suggested the term Narrative Learning Objects for this type of smart game objects and is developing the concept in her PhD thesis.

The application prepares the game objects for the demands of story generation to create a multi-agent system of narratively intelligent game objects. Game objects are imbued with additional information from three domains: game world common sense: drama tags and learning content.

Game World Common Sense. Game objects carry the common sense knowledge associated with their standard (and some conceivable non-standard) use, including common sense knowledge about their usual environment and interaction partners. Thalmann and Kallmann [8] developed a similar approach for 'smart objects' that carry a definition of their interaction features with virtual humans, including their functionalities, purposes and scripts.

Drama Tags. The story generation system appraises the game objects - including the procedures they are part of - according to their ability to achieve a specific emotional effect upon the character - generally triggering a character's hopes or fears, helping him towards his goal or pushing him towards failure. Dramatic appraisals will be altered according to changing, individual emotional preferences and fears of the character. While many game objects will retain their dramatic appraisal from previous game episodes (a fire or a gun will achieve the same dramatic effect on most characters), other dramatic appraisals will be altered according to changing, individual emotional preferences of the character (such as a dislike of raisins or preference for daffodils). Given its repertoire of dramatic functions for the character, every game object can now autonomously find its place in a dramatic scene – a scene that will be emotionally effective for the character.

Learning Content. In a story context, players have a high motivation to overcome an obstacle in order to reach their goal. This (artificially created) motivation can be used to initiate learning episodes. An obstacle-object can integrate any learning task such as a fact-related question or a complex procedure (A wants B, faces obstacle C that he can only overcome by solving task D). Every learning object is comprised of one 'success' script – a procedure to successfully overcome the obstacle – and many 'failure' scripts. Learners have to complete the procedure correctly in order to continue on their quest. As specified by Roger Schank in his model of Goal-based Scenarios, the learners are pursuing goals of their own choice and are ready to explore, investigate and learn in order to overcome their obstacles. More importantly, they are ready to risk failure while attempting to solve their tasks. [17] Learning and exercising knowledge then becomes a natural part of the story as the player faces obstacles and overcomes them. Different types of knowledge can be integrated into the setting. The right path to overcome an obstacle could be a correct answer (e.g. a fact, a socially sensitive reply), or performing the right procedure (e.g. taking part in a medieval ceremony, repairing a car).

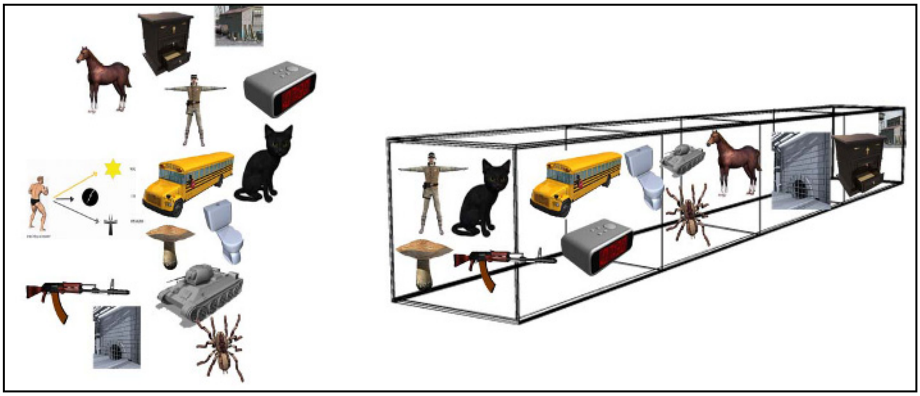


Fig. 7. The dramatic appraisal of game objects is translated to a sequence of encounters

Synthetic actors comprise a special class of game objects. Just as with other game objects, non-player characters and their AI will be structured around how they can dramatically become an obstacle or a supportive force for the protagonist, how they can initiate learning episodes, and what - common sense - interaction features they exhibit in the game environment. A synthetic actor working as a waiter in a bar might be dramatically capable of intimidating or supporting the characters, test learners with the social roles and scripts of his environment and enrich the storytelling of his scene with ontological, physical details about his particular game environment. As models for socially intelligent agents and virtual humans develop, our application will offer a chance to test just how believable these agents can talk about themselves as humans – and give the story generation system another slice of world knowledge to work with.

4 Generating the Game

In order to allow for interactivity while maintaining a dramatic story, our applications attempts to rebuild the world around the players with every interaction step to enable an ideal succession of dramatic encounters. This approach to storytelling maintains that fictional characters do not move in fixed, pre-existing world, as many world-based story generation systems assume. They move in a world that is bent and moulded around them by the storyteller, catering to the demands of dramatic pacing.

To a large extent, storytellers need to work in logistics, moving characters and objects around on a map. In its basic form, this process translates to tracking game objects and their possible trajectories on a map. In its more advanced form, the process is about enabling and planning encounters of the main character with game objects *at any place, time and frequency* desired by the plot. Generally, game objects need to be rushed towards an encounter with players, or vice versa, players need to be moved into encounters.

This interactive rebuilding of the world is limited by the player's knowledge of the world. E.g. as soon as the player has entered a street, the buildings of that street cannot be reshuffled or moved anymore. As soon as the player has met a game character, its movements are restricted by the laws of probability. Thus, the players evolving mental representation of the game world 'solidifies' the story world and makes it increasingly harder on the storyteller to force dramatically appropriate encounters.

This calls for an additional dramatic appraisal of game objects that is primarily concerned with the movement of players, effectively creating 'traffic direction' objects. While players may want to reach a certain destination, these obstacle objects will intercept them and redirect their movement more or less forcefully towards other locations and encounters. Just as game characters can be rushed towards the players in a coach, players can be whiffed towards encounters in subterranean sewer systems, mud slides or roadside kidnappings.

Another way of jumping creases in the map that is highly popular in films demands some advance planning. Game objects wishing to appear outside of their probability range may suggest previous scenes or information dropping ('planting') to create additional world rules that will justify their appearance: the gun could have been found by the bully character in a previous scene; the wife character usually lives in another part of town but could announce a visit to her parents in a previous scene. Game objects will therefore not only possess knowledge about their standard environment and interaction partners, but also some 'doorways' into alien game environments, with procedures how to access them.

5 Evaluation and Conclusion

We presented two approaches to work around the lack of world knowledge stifling the development of story generation systems. Both rely on a character-centred drama model developed by Frank Daniel that defines a story world emotionally through the eyes of a character. We hoped to show that a database of film scenes annotated in Frank Daniel's description format can offer a sufficient substitute for a story generation system's lacking theory of mind. Second, we suggested a way to integrate this character-centred drama model in a game world. Every game object is imbued with knowledge how to achieve a dramatic effect on the protagonist in its specific world environment (e.g. a door knows how to impede the protagonist). Finally, we showed how this approach can maintain a dramatic narrative regardless of the player's interactions. The system plans for encounters with game objects creating the appropriate dramatic effect, and rebuilds the game world accordingly. Combining these two approaches, we hope to create a system that is capable of generating individual game stories for a role-playing character in a large variety of game worlds.

Still, many questions remain unsolved. The use of annotated film scenes in a database raises questions about copyright. Also, if game objects and characters re-enact film scenes, albeit in new combinations, does this recombination of scenes deserve to be called story generation? A fundamental question concerns the interface from video-based story layout to animation generation in the game world. Can the annotations contain a higher-level description of the character's or object's goals and strategies in the scene? Game objects and characters will need to perform their dramatic effects on the protagonist in varying contexts, as their encounters with the protagonist are moved into the player's unpredictable trajectory. Further research is needed in two areas that will decide over the value of the proposed architecture: how can the scene annotations describe the dramatic functions, goals and strategies of objects in varying levels of context-dependency? And: How can these directions be understood as high-level commands by autonomous game objects and characters?

References

[1] Aylett, Ruth S.: Narrative in virtual environments - towards emergent narrative. AAAI Fall Symposium 1999

[2] Bailey, Paul: Searching for Storiness: Story-Generation from a Reader's Perspective. In: Mateas, Michael and Phoebe Sengers: Narrative Intelligence. Papers from the AAAI Fall Symposium, North Falmouth, Massachusetts 1999. (Technical Report FS-99-01) AAAI Press 1999, pp. 157-164

[3] Bringsjord, Selmer and David A. Ferrucci: Artificial Intelligence and Literary Creativity: Inside the Mind of BRUTUS, a Storytelling machine. Hillsdale, NJ: Lawrence Erlbaum 2000

[4] Cavazza, Marc; Charles, Fred; Mead, Steven J.: Emergent Situations in Interactive Storytelling. In: Proceedings of the 2002 ACM symposium on Applied Computing, Madrid, Spain, 2002, pp. 1080 - 1085

[5] Horton, Andrew: Writing the character-centred screenplay. Berkeley, Los Angeles: University of California Press 1994

[6] Howard, David and Edward Mabley: The Tools of Screenwriting. A Writer's Guide to the Craft and Elements of a Screenplay. New York: St. Martin's Press 1993

[7] Iuppa, N., Weltman, G., and Gordon, A.: Bringing Hollywood Storytelling Techniques to Branching Storylines for Training Applications, Proceedings of the Third International Conference for Narrative and Interactive Learning Environments, Edinburgh, Scotland 2004

[8] Kallmann, Marcelo and Daniel Thalmann: Modeling Behaviors of Interactive Objects for Virtual Reality Applications. In: Journal of Visual Languages and Computing, 13(2) 2002, pp. 177-195

[9] Laurel, Brenda: Towards the Design of a Computer-Based Interactive Fantasy System. Ph.D Diss., The Ohio State University 1986

[10] Mateas, Michael: Interactive Drama, Art and Artificial Intelligence. Ph.D. thesis. Technical Report CMU-CS-02-206. School of Computer Science, Carnegie Mellon University 2002

[11] Meehan, J. R.: The Metanovel: Writing Stories by Computer. Ph.D. Dissertation, Dept. of Computer Science, Yale University 1976

[12] Okada, N., and Endo, T.: Story Generation Based on Dynamics of the Mind. Computational Intelligence 8, 1992, pp. 123–160.

[13] Riedl, Mark Owen and R. Michael Young: An Intent-Driven Planner for Multi-Agent Story Generation. In: Proceedings of the 3rd International Conference on Autonmous Agents and Multi Agents Systems 2004
[14] Schank, Roger C.: Tell Me A Story. Narrative and Intelligence. New York: Scribner 1990
[15] Schank, Roger: Dynamic Memory Revisited. Cambridge: Cambridge University Press 1999
[16] Schank, Roger C., Abelson, Robert P.: Knowledge and Memory: The Real Story. In: Wyer, Robert S. (Ed.): Knowledge and Memory: The Real Story. Hillsdale: Lawrence Erlbaum Associates 1995, pp. 1-85
[17] Schank, Roger; Neaman, Adam: Motivation and Failure in Educational Simulation Design. In: Forbus, K.D. and Feltovich, P. (eds.): Smart Machines in Education. The coming Revolution in Educational Technology. London: AAAI Press / MIT Press 2001, pp. 37-69
[18] Schneider, Oliver; Braun, Norbert; Habinger, Gregor: Storylining Suspense: An Authoring Environment for Structuring Non-Linear Interactive Narratives. In: Skala, Vaclav (Ed.): European Association for Computer Graphics (Eurographics): Journal of WSCG Volume 11 No. 3, 2003. Plzen : University of West Bohemia 2003, pp. 411-417
[19] Seger, Linda: Creating Unforgettable Characters. New York: Henry Holt 1990
[20] Phoebe Sengers: Narrative Intelligence. Papers from the AAAI Fall Symposium, North Falmouth, Massachusetts 1999. (Technical Report FS-99-01) AAAI Press 1999, pp. 157-164
[21] Sgouros, N. M.; Papakonstantinou, G.; and Tsanakas, P.: A Framework for Plot Control in Interactive Story Systems. Technical report, Dept. of Electrical and Computer Engineering, National Technical University of Athens 1996
[22] Turner, Scott R.: The Creative Process: A Computer Model of Storytelling and Creativity. Lawrence Erlbaum Associates 1994
[23] Weyhrauch, P.: Guiding Interactive Drama. Ph.D. Thesis, School of Computer Science, Carnegie Mellon

How Do We Build This Thing?: Imagining Frameworks for Personal Narratives

Armele Adams

Culture Lab
University of Newcastle upon Tyne, UK
a.v.adams@newcastle.ac.uk

Abstract. In an increasingly fragmented world people's need to keep and communicate a coherent life narrative is vital. The paper takes as its starting point material presented at Culture Lab's *Personal Narratives* seminar in March 2005, as well as other popular forms of personal narrative and discusses their differing structures.

1 Introduction

Digital tools that can be readily used by a broad public are fostering new kinds of stories, to be enjoyed in the intimacy of home environments or in the potentially vast fora offered by the Internet. The spread of instruments that record and edit sounds and images, and use of graphics and word processing to generate pictures and texts, is providing a mushrooming population of user / authors with the means to recount its own stories. The traditional storytelling repertory of photo albums, slide shows and home movies is today extended by blogs, Quicktime files, and customised audiovisual features for a wide array of fixed and mobile computing devices.

The social backdrop against which personal stories are being created is equally strongly influenced by the upsurge in new tools: webcams and surveillance cameras behind televised reality shows are part of everyday urban and online iconography. Soap opera glamour of showbiz heroes is increasingly being offset not only by mediatised figments of their private lives, but also by the self-exhibitions of countless seemingly banal individuals investing the public spaces of networks.

These new kinds of distribution, networking and interactions of authorial power point towards potential possibilities of a participatory, expressive citizenship, notably within the European Information Society Technologies R&D programme, Since the late nineties, activities geared towards "generating creative content through (…) real-time authoring and design systems and skills development", and "managing digital content by supporting distributed and networked content", feature as research priorities for EU Framework programmes. [1] Regardless of how close such visions of networked-enhanced expressivity are to actual emerging realities, changes in our socio-technological environment undeniably affect how and why we tell and share our stories.

In response to the above, on the 18th March 2005 Culture Lab at the University of Newcastle upon Tyne hosted an interdisciplinary seminar under the heading *Personal Narratives*, drawing on artistic and sociological perspectives to feed discourse and de-

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 199 – 208, 2005.

fine key similarities and differences in this area. [2] Consisting of short presentations from multiple disciplinary perspectives, it focused on the use of electronic, interactive and mixed media to construct personal narratives. Participants were drawn from the University's research community and beyond. Questions of public profiling and social and creative positioning were discussed with respect to building portfolios and resumes, (pseudo-)autobiographical video documentaries, and art works employing self-portraiture including creative writing, live performance, sound constructions and installations.

The reflections that follow are grounded in and triggered by this seminar discussion and aim to identify commonalities of structure which can be applied to the process of constructing the personal narrative using a number of different media.

2 Why Do We Need Personal Narratives?

In *The River* Bruce Springsteen sang:

> I come from down in the valley
> Where mister when you're young
> They bring you up to do like your daddy done [3]

The stylised story contained within the song embodies the nostalgic feel of some late 1970s and early 1980s music for a life narrative that was already written and that one had merely to follow to its logical conclusion. This was mapped out in advance by parents, teachers and members of the community and involved a job (for life), maybe marriage and children and a secure place within a settled order. However, increased fragmentation within the workplace as companies merge, go under and outsource core functions mean that the continuity that once marked many life experiences is no longer attainable, although it may remain as a nostalgic ideal, with workers forced to move job, location and even shift through a number of different fields throughout the average career. In his analysis of the personal, emotional and social effects that these changes have had on specific individuals Richard Sennett comments that:

> The short-term, flexible time of the new capitalism seems to preclude making a sustained narrative out of one's labors, and so a career. Yet to fail to wrest some sense of continuity and purpose out of these conditions would be to literally fail ourselves. [4]

Anthony Giddens argues that personal narrative in itself is vital to human well being:

> The existential question of self-identity is bound up with the fragile nature of the biography which the individual 'supplies' about herself. A person's identity is not to be found in behaviour, nor - important though this is - in the reactions of others, but in the capacity *to keep a particular narrative going*. (…) It must continually integrate events which occur in the external world, and sort them into the ongoing 'story' about the self. [5]

This need for meaningful personal stories may be made all the more urgent in response to awareness of the demographic and transactional profiles that are constantly being compiled by clickstream data and inference engines that monitor our online activities, store cards that allow merchants to record and predict our spending habits and widespread moves to incorporate personal information within smart chips on passports and identity cards.

In his seminar presentation [6] business informatics researcher James Cornford's analysis of the use of the individual's story with regard to the creation of e-portfolios proposed a threefold model of the personal narrative (see below).

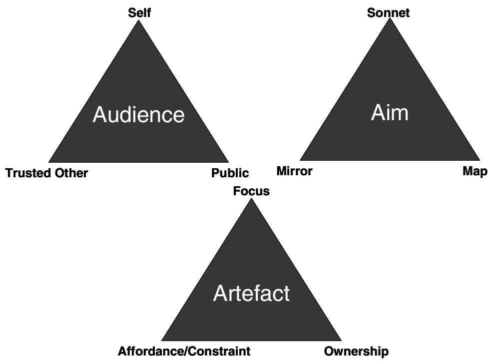


Fig. 1. Cornford's model of the e-portfolio

But how can we use these definitions to describe the personal narrative as a specific form? Although Cornford's model focuses on three aspects of personal narrative (Audience, Aim, Artefact) for the purposes of simplifying this (subjective) reading we have chosen to use only one, the AIM aspect, taken from Diez's analysis of the professional portfolio. [7] Diez separates the process of creating the portfolio into three: **Mirror . Map . Sonnet**. We propose to further discuss these metaphors, as we understand them, and to use the structure as a basis from which to describe and comment on the differing structures of some of the varied kinds of narrative presented both at the Personal Narratives seminar and elsewhere.

2.1 The Mirror

The metaphor of the Mirror suggests the unmediated collecting of information, objects reflected without selection or emphasis. Examples of unstructured outpouring of information can be recognised within a number of popular forms: the weblog at its most stream-of-conscious, the 'always on' webcam, and the annual 'Big Brother' live footage where the audience is invited to observe the housemates go about their daily routines, and even (as in Warhol's 1963 film) to watch them sleep.

In terms of personal narrative the Mirror stage might involve the collection and analysis of the whole 'story', unmediated by notions of commerce or value but with immense value to the self and to society. David Butler of Newcastle University's Fine Art Department described how, in creating *Spare Time Job Centre* [8] a process-based installation at the Chisenhale and Globe Galleries, the artist Ella Gibbs invited visitors to find out what skills they had developed in their spare time that could be offered to others within the community, either traded or given for free. [9] The work aped the styling of the UK Job Centre, (a government run space where the unemployed are asked to examine their employability in market terms) but instead asked participants to fill in forms which asked who they 'really were'. Talents such as playing the guitar, or, more unusually 'How to get something for nothing' were revealed, sometimes for the first time, as valuable to others, and passed on within a space that encouraged participants to enter into an invited social discourse about what was of value within their own particular experience.

Barrie Stephenson subsequently spoke about his experience on the BBC's *Telling Lives* digital storytelling project [10], involving week-long workshops during which participants focused on a personal story. [11] He described how each workshop would start with the group (at this point total strangers to each other) in a 'Story Circle', within which they would recount and explore their material. One woman brought her late mother's handbag (kept, unopened, for eighteen years) and slowly unpacked it in front of the group, reflecting on each object as it emerged. The experience was powerful and moving for all involved, a mirror held to her own personal story.

This 'outputting' of personal experience, in order to reflect on it, seems to be a powerful human urge: participants in the Personal Narratives seminar recounted experiences of situations where total strangers would habitually unburden themselves. One, while working as a TV-repair man, often learnt highly personal details about the lives of the people whose homes he was visiting. Another described the elevator in the twenty-one storey apartment block in which her parents lived, in which she had experienced a number of people passing on the intimate details of their lives, sometimes banal, sometimes scandalous. She only wished her parents lived higher than the seventh floor as she often missed the end of the story!

2.2 The Map

Just as this process of mirroring, of uncovering and questioning what exists, is essential to human experience, then also in order for the resultant profusion of personal information to convey something other than raw data overload, storytelling patterns appropriate to generating new mechanisms for mapping and navigating, identification and recognition are being co-invented by authors and audiences.

A common example of authorial mapping is the ubiquitous photograph album, present in most households, within which photographs are sorted and presented in a way that is meaningful to the creator.

Julian Germain's recent exhibition of photographs of Charles Albert Lucien Snelling as well as blown up pages from the photo albums Snelling had created over many years [12] foregrounds and demonstrates this incremental process of self actualisation, the gradual accruing and organising of snapshots intended for an audience of the self or close contacts who already have some knowledge of the person and family group. Commonly the photograph album is experienced either alone or forms the framework for an improvised story about the subjects pictured within.

Apple has recognised and attempted to satisfy this urge towards narrative order, offering the functionality within their iPhoto tool of creating virtual 'albums' within which images can be organised using different criteria, animating these within an on-screen slideshow, and even offering the service, in conjunction with Kodak, of ordering a professionally printed, hardcover book of an album, as well as web publishing capabilities. Images can be repurposed to appear within a number of different albums, allowing users tell an infinite number of stories with the same material. On its website Apple urges users to:

> Set up baby-to-adult albums at a wedding reception; Share your vacation with family or friends on iPod or send an iDVD; Take photos of your class during the year and create a slideshow to share with your students at year's end; Take friends and family on the East coast on a walking tour of your new West coast digs; Create a photo essay for your social studies class [13]

Equally through the work of professional street and portrait photographers from Eugene Atget, James Van Der Zee and James P Ball through to Richard Billingham and Nick Waplington in the present day we are able to access wider stories of communities that are either no longer in existence or to which we have no access.

Fig. 2. Images by Ball, Van Der Zee and Atget from C19-C20

2.3 The Sonnet

These collections of photographs, amateur and professional, tend to record meaningful events in the lives of their subjects: holidays, weddings, christenings, parties of all sorts, times when people come together or feel further apart. These are the nodes or landmarks around which personal lives are structured and navigated. Giddens observes that:

> Personal calendars are timing devices for significant events within the life of the individual, inserting such events within a personalized chronology. (..) 'When I got married,' as a basic date within a life-plan calendar (..) may be largely ousted by 'when the marriage broke up' as a more significant psychological marker. [14]

Just as people make use of these 'life landmarks' around which to structure their own life narratives, altering the order or discarding unwanted items to better shape the story, so too conventional fiction uses structure to create stories that will hold the audience's attention and stimulate an emotional response. Margaret Wilkinson's seminar presentation focused on her adaptation of Syd Field's scriptwriting approach *Classic Plot Design*. [15] A plot has common elements, it has shape (like the photograph album) but crucially it also builds and has causality. This is created by examining the meaningful events within the narrative and deciding which ones to emphasise and which to de-emphasise in order to create an arc of experience, dramatic tension and, crucially, to reach for an emotional rather than a historical truth.

The rules of *Classic Plot Design* assist the creator in pacing the story in a way that meaning is implied inherently within the structure of the piece. This becomes necessary once the retelling of the narrative becomes separated from the face-to-face improvisation inherent in the use of the photograph album. By finding a trigger (a single incident that represents a change from the status quo) and three plot points (important incidents of **rising intensity**) writers can avoid their story getting lost in incident and will sustain audience interest throughout. The idea, in Wilkinson's words, is:

> .. not to dramatise everything. Differentiate between crucial scenes (plot points), less important scenes, and bridging scenes that are communicated in summary or narration. [16]

The final, but most important, part of this narrative structure is the Narrative Question, arising at the beginning of the piece. This is 'a central issue, posed as a question, the one thing you want the reader to wonder about, worry about, throughout' [17] and which is resolved somewhere near the end of the story.

The BBC's *Telling Lives* project approached the coherent communication of individuals' personal landmarks in a similar way, providing participants with a standard structure on which to map the story that had previously been reflected upon in the earlier phase of the workshop, encouraging them to discard some elements and focus on others in order to produce their own 250-word script to be read out over 12-20 still images presented over two minutes. Barrie Stephenson explained that within a limited creative timeframe this structure enabled people to focus in on the elements vital to communicating their own life stories.

The structure offered to these debut filmmakers can be likened to the sonnet form: a strict formula that allows the poet infinite variation and improvisation within its own constraints. It also created a 'safe' structure within which inexperienced filmmakers could successfully recount a basic narrative, giving confidence to subsequently tackle more complex creative challenges.

The films were created to be broadcast in isolation of each other between longer shows, but on archiving these pieces on the Web the BBC chose to group the pieces in ways that create linkages in the minds of visitors to the site, with ordering, for example, by geographical location. At this point the project becomes more interactive for experiencers, allowing visitors to the site to create their own narratives by choosing, rejecting, ordering their own viewing of the films by personal preference. This kind of structure allows individuals to map their own Personal Narrative but also allows others to plot a course through them.

While the ordering of films on the *Telling Lives* website allows interaction tailored to the specific interests of the experiencer, it does not allow for the building of a narrative arc which will necessarily sustain interest and elicit emotion from the

Fig. 3. Richard Billingham 'Untitled' 1994 & 1996 courtesy of Anthony Reynolds Gallery, image from Nick Waplington 'The Wedding' 1996 reproduced by kind permission

Fig. 4. Eugene Atget, rue du Maure c1908

experiencer. The Sonnet structure employed within each short film benefits from its constraints by offering a more 'complete' emotional experience.

Similarly the *Spare Time Job Centre*, a piece of process based art concentrating on the stimulation of discourse amongst participants, was also presented as an installation in order to communicate the methods and aims to a wider audience than those who had taken part. Elements were chosen for display, others discarded in order to enable the creator to guide the experiencer towards her own emotional truth, arrived at through a process of mirroring and mapping. At this point the story becomes truly 'artful' and the experiencer can appreciate and be entertained by the product in a way that is useful to her:

> The reader of a novel, the spectator at a play, experiences the particular comfort of seeing people and events fit into a pattern (..) the "moral" of narrative lies in the form .. [18]

At this point the narrative becomes a commodity, a discrete object that can be passed on without the creator needing to be present for its enjoyment. It is complete.

2.4 Mapping the Sonnet

These metaphors are relatively easy to apply to areas where the artistic product, once created, does not change. However there is an element of liminality within live performance, highlighted by two contributions to the seminar, which challenges this.

In his monologue *No Small Inheritance*, created as part of 'Testament', a personal stories project recently created and performed as part of the RSC season at Live Theatre, Newcastle, Tom Shakespeare uses life writing as a tool to communicate issues of genetics and disability within a narrative framework which engages people in an emotional and entertaining rather than purely intellectual way [19]. The process involved collecting diverse personal recollections, reflection, mapping and finally structuring the material before performing it in front of an audience, when it may give rise to further personal stories from the audience [20]. He will be developing this way of working further with the support of a three year NESTA Fellowship [21].

Will Edmondes, an artist working with sampled and improvised sound, described how he works to a loose structure previously decided on before a performance, which enables structuring of a coherent piece of work [22]. This Sonnet exists only while it is being experienced during a performance but it requires the use of the Map to facilitate its improvised creation. Similarly Joel Ryan has recently discussed useful new ways of mapping digital music scores and instrument layouts within the experimental electronic music field [23]. The nature of the method of working allows for subtle and not so subtle reactions to audience and environmental input while keeping to a predetermined intention.

If, as we have stated before, the Map structure allows for successful interaction, and the Sonnet for an emotionally satisfying experience, then this would suggest that live performance is an area where both can be attained.

3 Next Steps

Diez's metaphor does seem useful in analysing the process and discrete elements of different kinds of personal narrative and to identify which of the three Mirror, Map, Sonnet elements are necessary in creating a successful narrative for different purposes, even allowing us to identify the potential for interaction plus emotional resonance which stems from the iterative nature of the creative process within performance.

It would seem that at the most basic (the unpacking of the handbag) only one of the three elements is necessary in creating some form of narrative, but that in order for the personal narrative to be enjoyed more than once all three elements must be passed through, either independently, one by one, or as part of a creative cycle.

The above text is intended to be used as the basis for further discussion with colleagues working across a wide range of disciplines, with the intention of suggesting and refining models for future work. A further event is planned for March 2006.

Acknowledgements

Thanks to David Butler, James Cornford, Will Edmondes, Tina Gharavi, Richard Grayson, Sally Jane Norman, Tom Shakespeare, Barrie Stephenson and Margaret Wilkinson for their participation on the day and further discussions, Julian Germain, Anthony Reynolds Gallery and Nick Waplington for permission to use images of their work in this document and the accompanying presentation. The Personal Narratives seminar was performed in the frame of the INSCAPE Integrated Project (EU RTD contract IST-2004-004150) which is funded by the European Commission under the sixth Framework Programme. More information on the project is available at www.inscapers.com

References

1. European Union. Information Society Technologies Programme (5th Framework). EITC 97 Secretariat. Creating a user-friendly information society: Working document Excerpted from COM(97) 553. 07 November 1997. 12/08/05 (http://www.cordis.lu/esprit/src/istwork.htm).
2. Personal Narratives. 18 March 2005. University of Newcastle upon Tyne, Newcastle. Full programme at www.ncl.ac.uk/niassh/culturelab/personal_narratives.htm. 12/08/05. Cf infra. acknowledgements.
3. Springsteen, Bruce: The River. By Bruce Springsteen. Rec. 1980. The River. Sony, 1980.
4. Sennett, Richard: The corrosion of character: the personal consequences of work in the new capitalism. New York: Norton, c1998. p122.
5. Giddens, Anthony: Modernity and self-identity: self and society in the late modern age. Cambridge, UK: Polity Press in association with Basil Blackwell, 1991. p54.
6. Cornford, James: Portfolios as vehicles of personal narratives. INSCAPE. Personal Narratives. Devonshire Building, University of Newcastle. 18 March 2005.
7. Diez, Mary: The Portfolio: Sonnet, Mirror and Map in Burke, Kay: Professional Portfolios. Arlington Heights: IRI Skylight, 1996. pp. 17-26.
8. Gibbs, Ella, comp. Spare Time Job Centre. 29 Jan. 2001. Chisenhale Gallery, London. 12/08/05 (http://www.chisenhale.org.uk/html/files/319_project_info.html).
9. Butler, David: Artists' personal narratives and social positioning. INSCAPE. Personal Narratives. Devonshire Building, University of Newcastle. 18 March 2005.
10. Stephenson, Barrie, comp. BBC - Telling Lives. 25 August 2004. BBC. 12/08/05 (http://www.bbc.co.uk/tellinglives).
11. Stephenson, Barrie: Telling Lives: authoring personal stories for television. INSCAPE. Personal Narratives. Devonshire Building, University of Newcastle. 18 March 2005.

12. For every minute you are angry, you lose sixty seconds of happiness: Baltic Centre for Contemporary Art, Gateshead, 05/03/05-05/06/05. 12/08/05 (www.balticmill.com).
13. Apple Computer Inc. Explore iLife, How can you use slideshows? 2005. 01/09/05 (http://www.apple.com/ilife/iphoto/slideshows.html).
14. 14.Giddens, Anthony: Modernity and self-identity: self and society in the late modern age. Cambridge, UK: Polity Press in association with Basil Blackwell, 1991. p85.
15. Field, Syd. Screenplay: The Foundations of Screenwriting. New York: Delacorte Press, 1979.
16. Wilkinson, Margaret: Autobiography in Three Acts. INSCAPE. Personal Narratives. Devonshire Building, University of Newcastle. 18 March 2005.
17. ibid
18. Sennett, Richard: The corrosion of character: the personal consequences of work in the new capitalism. New York: Norton, c1998. p135.
19. Tom Shakespeare has played an important role as the Director of Outreach at the Policy, Ethics and Life Sciences Research Institute (PEALS) in Newcastle, notably opening up public debate and stimulating wider engagement in controversial areas of life science development. 01/09/05 (http://www.ncl.ac.uk/peals).
20. Shakespeare, Tom: No Small Inheritance. INSCAPE. Personal Narratives. Devonshire Building, University of Newcastle. 18 March 2005.
21. NESTA. Our Awardees: Tom Shakespeare Profile. 2005. 12/05/05 (http://www.nesta.org.uk/ourawardees/profiles/5017/02_profile.html)
22. Edmondes, Will: Diary/sketchbook-dictaphone personal narrative improvisation. INSCAPE. Personal Narratives. Devonshire Building, University of Newcastle. 18 March 2005.
23. Ryan, Joel: MuViz: Prospecting for New Composition and Performance Tools for Electronic Music in Brouwer, Mulder & Charlton: Making Art out of Databases. Rotterdam: V2_Publishing/NAi Publishers, 2003.

Beneficial Dependencies: Design Principles for Narrative Games

Ido A. Iurgel

ZGDV, (Computer Graphics Center), Fraunhoferstr. 5,
Digital Storytelling Group,
64283 Darmstadt, Germany
Ido.Iurgel@zgdv.de
http://www.zgdv.de/zgdv/departments/z5

Abstract. This paper presents the leitmotifs of current research on psychologically colored interactive narration. Virtual Characters, their stories and adventures are in focus, and not the story of the user. The concept of a Capricious Non-Player Character (C-NPC) is introduced as a new kind of agent appropriate for such stories. A C-NPC expects consideration and support, even towards its technological limitations. A narrative educational game is presented that functions as a testbed for the creation of C-NPCs. An overview of the technology employed for dialogue management is provided.

1 Introduction

How can we exploit the possibilities of identification and empathy with a virtual actor, and its intense emotions towards situations in the story, in an interactive drama? This paper presents ongoing R&D towards increased engagement with virtual characters and *their* drama. The paper provides an overview and justification of the concepts that guide this work, and of current research efforts.

Some of the leitmotifs of this work – justifications follow in the next paragraph – are these:

- It is not necessary and most difficult to assign the main role to the user. A dramaturgically less pronounced role is often better. An excellent role for the user is that of the side-kick, i.e., friend and assistant of a virtual protagonist.
- Interactive drama and games are not contradictions. Certain advantages exist when defining clear goals and the possibility of failure.
- Use natural language input and output, but do not pretend to simulate real conversations. The user should adapt to the limited capabilities of the virtual characters.
- Equally, the user should adapt to the psychological peculiarities of a virtual character, which does not pretend to simulate realistically human psychology.
- Provide the user with a good reason to adapt to the limitations of technology, e.g., this would be the only way of winning the narrative game.
- A story is usually about a person trying to cope with a difficult situation. Exploit psychological aspects of this coping, as part of the game challenge. Take special care of inner transformations of the virtual main person.

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 209 – 218, 2005.

- Consider, in your narrative game design, the creation of a mutual dependency of player and virtual character. E.g., the user needs the character to win the game, and the character needs the user to survive or to be happy.
- If you are using natural language input, be funny or weird in the design of you story; do not attempt to be tragic.

The next paragraph elaborates on the aforementioned principles. Then, an educational game is presented as a study on how to create appropriate virtual characters. They will be called “Capricious Non-Player Characters”.

2 On Being a Side-Kick of a Capricious Non-player Character

Usually, in traditional narration, e.g. cinema, the spectator is touched by the emotions of other people, e.g. of the protagonist. The most fundamental leitmotif of the research described here consists in exploiting the notions of identification with the situation of others, of psychological understanding and empathy, in order to create interactive drama. This notion is derived from the psychology of traditional narration.

A straightforward and simple example of the psychological narrative envisaged requires a virtual protagonist in need of the user to cope with the difficult situation it is in. The user, playing the role of a side kick (or another supportive role), advises, consoles, accompanies and explains. The player wins the game e.g. if he guides the protagonist to the desired happy end. Many more complex variations of this theme are conceivable.

When more weight is laid on psychology than on action, the inner transformations and conflicts of the protagonist move into focus. The player would assist the protagonist with these transformations. That’s where employing natural language becomes important, since psychological exploration and support of the protagonist would need elaborated communication with it. Taking into account the current stage of natural language technology and psychological, communicative and social modeling, the narrative game design must incorporate the involuntary comic of technological shortcomings and failures. That’s why a funny or weird story environment and characters are necessary.

Psychology and communication of the virtual protagonist won’t be realistic. Hence, the user has to adapt, and he should be given a good reason to do so by enforcing this adaptation as part of the game challenge. This is the most important motivation to design the application as a game with challenges and the possibility of failure – turning shortcomings of the system into game challenges.

The R&D step currently being undertaken towards this psychologically colored narrative game focuses on the design and implementation of the virtual actors. What are the concepts, methods and technologies that will lead to highly interactive, emotionally engaging virtual actors for interactive stories? Our current efforts concentrate on Capricious Non-Player Characters (C-NPCs). A C-NPC is a virtual character that does not pretend to be psychologically and linguistically realistic, but that expects to be supported and endeared, and that demands a considerate treatment towards its (technologically caused) linguistic, psychological and social limitations and quirks. It is capricious in the sense that it will reward consideration, but blame the user for lack of adaptation, possibly even abandoning the user as a consequence. A virtual protagonist in need of a human friend would be a C-NPC.

The testbed for the development of C-NPCs is currently the educational game *Ask & Answer*. This game and the role of the C-NPC in it are explained in paragraph 4.

3 Related Work

The creation of social bounds with intelligent virtual agents was examined e.g. by Cassell and Bickmore [1], with focus on small talk and increase in credibility through it. Prendinger and Ishizuka [2], [3] have developed emotional agents and scripting languages for believable virtual agents with autonomous faculties that can be authored. Paiva et al. [4] have studied the context of empathy with agents and bullying, presenting concepts that would allow for the creation of emphatic relations. Stonks et al. [5] have elaborated on the prerequisites that would be necessary for the establishment of a friendship between a human and a virtual human. These works did not address social bounds and emotion from the game oriented, strategic point of view addressed in this paper.

The effects of virtual characters on learning have been widely studied; a final conclusion is not yet possible. Vassileva's and Okonkwo's [6] studies suggest that the effect on the direct influence of learning efficiency is small, but that effects on motivation and enjoyment are very clear. A fierce advocate of games as the most natural way of learning for young "digital natives" is Marc Prensky [7].

Chris Crawford [8] has already extensively elaborated on the idea that it is up to the user to learn the language of the virtual character, and not the other way round, in interactive storytelling. A description of the special requirements of domain oriented dialogue management is provided by Bernsen and Dybkjær [13].

4 C-NPCs in *Ask and Answer*

Ask & Answer is a game and story on its own and could be extended into a larger narrative educational game. It is also an appropriate testbed for the creation of C-NPCs, because the game settings motivate the users to stay tuned to the peculiarities of the C-NPC, without demanding yet a full elaboration of the research agenda described above.

Ask & Answer targets pupils aged 10-15. The story that drives *Ask & Answer* goes as follows: Fritz, the main person and C-NPC, was the janitor of a now abandoned school, and, with the years, learnt everything that was taught at this school. It was quit by its wife when the story begins, and is in an instable emotional mood. Its wife reproaches it of only being interested in and speaking of school lessons and former teachers. At the beginning of the story, it denies this allegation; but, in the course of the narration and of reflections, it recognizes that there is some truth to it, and decides to try to fix the relation to its wife and to change its spots. The story ends with this decision.

Fritz allegedly needs the players to reflect about the dramatic situation (i.e. about its marriage and its behavior), to console and advice it, and to speak about the beloved former school lessons. It is through this chat with the players and through reflection on it that Fritz will recognize that it is fixed on this part of the school's past and not attentive enough to other persons and to its wife.

The love affair of the grumpy school janitor should appear ridiculous to pupils, and even more so their role as advisers and its need of consolation. This situation shall be funny and not tragic at all.

The dramatic role of the players consists in accompanying Fritz through this inner transformation. There are many possibilities to fail, in this scenario, when chatting with Fritz, and to displease the C-NPC: Not speaking about the former lessons when it wants to, not consoling it when it needs it, giving the wrong advices, being too rude or direct, using a language that it does not understand, and more.

Apart from Fritz' story, *Ask & Answer* is an educational, competitive game. It adopts the formats of some TV-shows for kids and families, where two teams compete and scores are given for correct answers. Two teams of 2-8 participants play over the network. A team scores when it knows the right answer to a question asked by the system. Each participant looks at his own screen and has its own speech enabled avatar. The exemplary current domains of learning are history and geography.

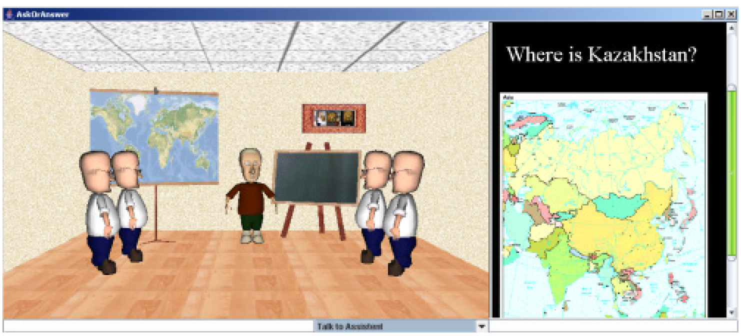


Fig. 1. Two teams confront each other, Fritz is in-between

Fritz knows in fact every answer to any question of the game. But it will only help if the player endears it enough and assists it with its story related problems. Fritz can only be heard by the team it is near to. If the team does not succeed in endearing the C-NPC, it might refrain from delivering the answer, or even worse, it might leave the neglecting team and change to the other team. But even if Fritz is satisfied with some group, it might first want to talk about some favorite topic (typically additional material related to the question), before giving away any solution.

This concept of a C-NPC much differs from the idea of a virtual character as believable actor or reliable assistant. Now, the virtual character must be understandable, rather than believable, so that the player may endear and manipulate it.

5 Technology

5.1 Framework of the Dialogue Management

The framework used for Dialogue Management (and behaviour management in general) is derived from research on combining different AI-technologies for interactive drama into a single system, with focus on intuitive and efficient *authoring* of these technologies. This research assumes that the development of the authoring methodologies for storytelling is as important as the choice of AI-technologies, and must accompany them.

Past experiences on creating interactive applications have shown that directed graphs are intuitive and perfectly suited for handling many interactive situations. In the context of dialogue management, graphs can be used e.g. for predictable dialogue moves like dialogue pairs as “Do you want to hear more about it?” → “yes/no/other”. The current challenge consists in providing the author with the means to smoothly change to and integrate symbolic control methods (e.g. rules, scripts, planning) when the graph approach attains its limits.

This combination of methods is achieved with a hierarchy of “frames”, certain nodes of a tree. This tree can be employed, in the basic case, as a hierarchical transition network, in the present application, as a State Chart without parallelism, but with strict encapsulation of states (i.e., no arrows to or from the inside of the composite state). A frame is then a composite state of a State Chart.

Additionally, a frame can be associated to a symbolic engine that may activate immediate daughter frames and leaves. In the current implementation, an instance of the production system Jess is attached to every frame, and the inner elements are also represented as Jess facts. I.e. not only transitions can activate a state, but also fireying rules. Every inner element (leaf or frame) of a frame, to which a production system was associated, can carry meta-tags which allow the rules of the production system to apply. The scope of these rules is restricted exactly to the states insides this frame.

With this system, a strict encapsulation of behaviour inside a frame is achieved, and the behaviour can be defined either with the rules system, with directed graphs, or with both, depending on the needs of the author. While rules in general cannot be represented graphically, at least important elements to which these rules are applied, the inner frames and states, have a graphical representation.

The processing steps of the system, in the case that an external event has to be handled, e.g. if the user has typed in something, starts at the currently active state. Assuming that an event has occurred, and that state A is active, and that A is immediately contained by frame F, the treatment is this:

1. If the conditions of an outgoing transition of A or of one of the enclosing frames apply, follow this transition. Else, go to 2.
2. Activate the production system associated to F. If the production system notifies that it could not handle the event, go to 3.
3. Activate the production system that belongs to the frame that contains F, and continue activating every subsequent father system, until a system handles the event, or until the top system is reached.

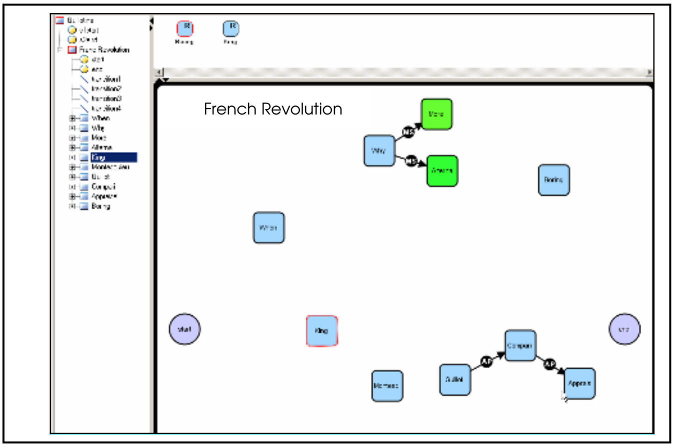


Fig. 2. Screenshot of the authoring tool "Cyranus". Note that the tree representation used in figure 3 is secondary for the author (narrow left window of Cyranus). The author will work *inside* a frame, where the direct daughters are represented. A certain set of rules attached to the frame can activate one of the daughters.

I.e., if no condition of a transition matches, delegate the event to the rule system of the immediate container frame, which might activate a contained element or delegate the event to be handled by its own father frame. This process assumes that every step will apply a less specific strategy, i.e. the best strategy is to follow a transition, if the event was predicted by the author, and otherwise to apply the rules of the nearest frame leading up in the hierarchy.

This framework was first developed within the art-E-fact project[1], and work is continued in context of the INSCAPE[2] and the Virtual Human[3] project. In the later, the extension with machine-learning technology is being implemented, which shall enable an interactive authoring process.

5.2 Dialogue Management

The aforementioned framework is being used for managing the linguistic and non-linguistic behavior of Fritz. The dialogue management shall enable

- Intuitive authoring of dialogue moves, enabling the author to create, intuitively, a personality with a peculiar language and themes
- Support pro-active behavior and mixed initiative dialogues
- Support domain/story oriented dialogue management, in contrast to task orientation.

The goal here is not as ambitious as to create "realistic" conversations, which is impossible with current technology. It suffices here that the results of the dialogue man-

---

[1] Cf. www.art-e-fact.org
[2] Cf. www.inscapers.com
[3] Cf. www.virtual-human.de

agement are understandable enough for a player to learn to maintain a successful conversation, but not as simplistic as not to require efforts and to allow for mistakes.

No standard methods exist yet for domain and story oriented dialogue management. Chatterbots are domain oriented, but usually not easy to author, do not support pro-active behaviour, and usually no serious dialogue state dependent behaviour is shown. The Dialogue Management of *Ask & Answer* follows the line of many chatterbots by employing canned text as output and trying to foresee possible user input, without neither employing nor formal syntactic or semantic analysis, nor a formal representation of the content. It is shallow in the sense that no deep linguistic analysis is made.

Typed text input of the user is handled with transitions or with rules. With transitions, the expected input is expressed as a Boolean condition, in praxis using an extension of the XEXPR functional XML-language. It is possible to express the expected input as a regular expression, or to express that any member of an equivalence class of inputs will fit. For this, ALICE[4] is employed as a module, in a non-standard manner, exploiting the "reduction" mechanism of ALICE to define equivalence classes of input. E.g. ALICE is used to define that "yes", "sure", "ok", etc. shall be treated as equivalent.

The encapsulation of the production system in the composite state (frame) allows creating different behaviors for different thematic frames, e.g. the virtual character may try first conciliations strategies if the user is insulting it, but may attempt at a broad exploration of the theme if it is chatting about the French Revolution with the user. When used within the scope of a rule set, a composite state may be viewed as an arbitrary hierarchical data structure that contains e.g. behavior directions for different situations. For example, a composite state C may be assigned to a certain question, a composite states D inside C may contain a set of utterances to be used when the user asks the question for the first, second, third, …, time, and another composite state E inside C may contain utterances that are employed to guide the user pro-actively to ask the right question.

When the transitions fail to handle text input, rules of the enclosing frame assume. This is made by attaching a regular expression or an ALICE reduction to a frame or simple state, as meta-data. This allows the rules to decide on whether a state shall be activated, due to the text input event. An example follows.

Fritz: "About the French revolution, I know everything! Ask me!"
Player: "When was the King killed?"
Fritz: "You've got me! I do not know that!"

In the example above, Fritz does not get immediately angry, in spite of not knowing the answer. This is part of its "personality". The rule set contained by this frame express that Fritz will get angry in case that it does not know an answer, *except* when it has prompted to ask. However, even then, its reaction changes if this happen too often. Thus, the player should normally try to foresee a question that Fritz will be able to answer. This would be normally about some fact that is in the curriculum of the school. Some more details of the system are described in [9].

---

[4] Cf. www.alicebot.org

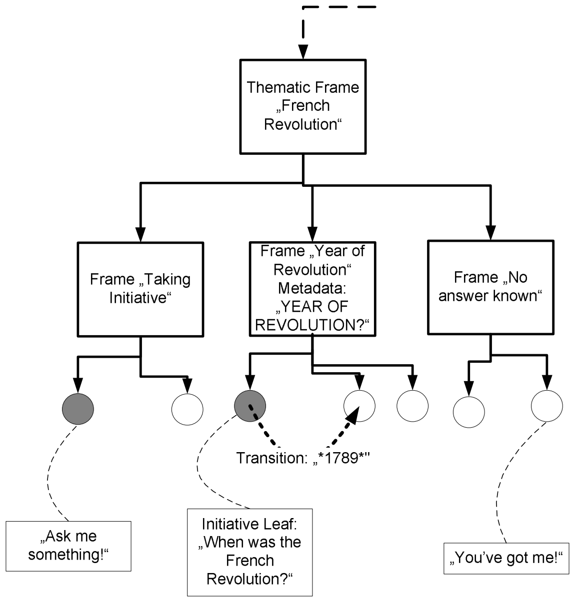


Fig. 3. Example content of a thematic frame

6 Future Work On *Ask and Answer*

Ask & Answer will be submitted to evaluation in the course of the year. It will probably later become a single chapter (or level) of a more complex story that exploit the ideas of creating social bounds and mutual dependencies within an educational, narrative game. Pragmatic considerations have so far restricted the implementation to a single “dramatic person”, other parts, e.g. of the antagonist, could have been assigned. This move is being considered.

7 Conclusion

This paper has presented the leitmotifs of current research on interactive drama based on identification, empathy and communicative virtual actors. The edutainment game *Ask & Answer* introduces the notion of Capricious Non-Player Characters, and explores this concept. In *Ask & Answer*, it is a matter of game strategy to understand and endear the C-NPC.

The ultimate goal consists in narrative games with psychologically elaborated C-NPCs. It is part of the game challenge to understand their inner conflicts, their social and emotional narrative situation, but also the technologically caused linguistic and behavioral peculiarities. Though fun and entertainment are in focus, the liaison of psychology, game and narration itself posses clear learning aspects: It is about learning to understand social constellations and psychological and emotional processes.

ELIZA was a bad psychiatrist, because it was not able to model the situation of the user, and it still lies beyond today's means to create a virtual psychologist that would support the user directly, by suggesting how to understand his psychological situation. But an inversion of the scenario should be at our fingertips: Creating a virtual *patient* the user has to support and understand psychologically. It is noteworthy how closely this scenario of an inversed ELIZA relates to the research on interactive narration sketched in this paper: The user would resemble a psychologist of a virtual actor, and he has attained the goal of the game when he has coped with and understood his virtual patient and its situation, i.e., *its story*.

This throws a new light on "interactive drama". It's yet impossible to create software that would advise on how to live an individual life and cope with life's problems. But it is possible to prepare typical *cases* and to develop interactive applications that entertainingly convey their messages.

A serious barrier is build by the knowledge acquisition bottleneck. Authors of linear stories work intuitively, and most of psychological or social psychological theory is not detailed and not formal enough to provide the necessary formal models. Hence the importance of the authoring methodologies, in order to support the creation of appropriate virtual characters and interactive stories.

Acknowledgements

This work has been partly funded by the Fifth and Sixth European Framework Programs, IST and FP6-IST, within the art-E-fact and INSCAPE projects, and by the German Federal Ministry of Education and Research, within the project Virtual Human.

References

1. Cassell, J., Bickmore. T.: Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. In: User Modeling and Adaptive Interfaces (2002), 12:1, 44
2. Prendinger,H., Ishizuka, M.: Designing and Evaluating Animated Agents as Social Actors. In: IEICE T.ransactions on Information and Systems, Vol.E86-D, No.8 (2003) 1378-1385
3. Prendinger,H., Ishizuka, M.: Evolving Social Relationships with Animate Characters. In: Proceedings of the AISB-02 Symposium on Animating Expressive Characters for Social Interactions (2002) 73-78
4. Paiva, A., Dias, J., Sobral, D., Aylett, R., Sobreperez, P., Woods, S., Zoll, C., Hall, L.: Caring for Agents and Agents that Care: Building Empathic Relations with Synthetic Agents. Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 1 (AAMAS'04) 07 19 - 07, New York City, New York (2004) 194-201
5. Stonks, B., Nijholt, A, van der Vet, P., Heylen, D.: Designing for Friendship: Becoming Friends with Your ECA. Workshop on Embodied conversational agents - let's specify and evaluate them! 16 July, 2002. Bologna, Italy. In conjunction with The First International Joint Conference on Autonomous Agents & Multi-Agent Systems 91-97

6. Vassileva, J., Okonkwo, C.: Affective Pedagogical Agents and User Persuasion. In: Proc. Universal Access in Human-Computer Interaction (UAHCI), held jointly with the 9th International Conference on Human-Computer Interaction. New Orleans, USA (2001) 397-401
7. Prensky, M.: Digital Game-Based Learning, Paragon House 2005
8. Crawford, Ch.: On Interactive Storytelling. New Riders, Berkley 2005.
9. Iurgel, I.: Narrative Dialogues for Educational Installations. NILE 2004, Edinburgh, 10th - 13th August (2004)
10. Mateas, M., Stern, A.: Natural Language Understanding in Façade: Surface-text Processing. TIDSE 2004, Darmstadt, Germany
11. Iurgel, I.: From Another Point of View: Art-E-Fact. In: TIDSE 2004, June, Darmstadt, Germany
12. Iurgel, I.: Virtual Actors in Interactivated Storytelling. IVA 2003, September, Kloster Irsee, Germany, 254-258
13. Bernsen, N. O., Dybkjær, L.: Domain-oriented conversation with H.C. Andersen. Proceedings of the Tutorial and Research Workshop on Affective Dialogue Systems (ADS04), LNAI 3068, Kloster Irsee, Germany, 14-16 June 2004, 142-153

Storytelling for the Small Screen: Authoring and Producing Reconfigurable Cinematic Narrative for Sit-Back Enjoyment

Ludvig Lohse and Maureen Thomas

Cambridge University Moving Image Studio
LL255@cam.ac.uk

Abstract. This paper describes the concept development and practical application of tools for editing and reconfiguring video and sound to create stories using broadband application. The material developed is *Gods in the Sky Choice*, a fully working experimental pilot demonstrator produced as part of the NM2 (New Media for the New Millennium) Integrated Project (IP) of the European 6th Framework Programme Priority 2 (Information Society Technologies) Call 2: Cross-media content for leisure and entertainment <http://www.ist-nm2.org/>

1 NM2 Production: Gods in the Sky Choice

Gods in the Sky Choice, derived from *Gods in the Sky,* an entertaining, factual three-episode documentary with high entertainment values produced by WagTV for Channel 4, first broadcast in the UK in August 2003 as three 50-minute programmes, tests conceptual thinking about structuring non-linear reconfigurable narrative for the screen by analysing and re-purposing the 150 minutes of footage for handling, via broadband, in three different modes: *Entertainment*, *Education* and *Information*, as a pilot demonstrator and model for future production.

The *Gods in the Sky Choice (GiSC)* practice-based research project has two main production goals:

- Using the BT prototype Flexible Media Toolkit (BT FMT), develop and produce a fully-working pilot demonstrator which handles 150 minutes of video with audio and delivers personalised versions of the content in 3 modes (*Entertainment*, *Education* and *Information*) with coherent narrative structure generated through real-time editing from the entire media pool for delivery and access via a PC and Windows Media Center (WMC).
- Develop and implement an effective, non-intrusive interactive framework for the interactive experience, using Adobe Photoshop, Macromedia Director, utilising ActiveX controls, based on the content of the programmes as well as their structure, for access via a WMC remote control unit.

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 219 – 222, 2005.

2 Process

For the experimental Pilot Demonstrator, material had already been shot and edited for the 3 x linear 50-minute broadcast episodes originally broadcast by Channel 4 in 2003, so the process was one of 'reverse engineering'.

The analysis of the existing material in terms of programme-structure, format and specific content - conducted in discussion between the editor/producer and the script consultant - used paper and computer-generated diagrams and schemata, and simultaneously employed the BT FMT to experiment practically with the digitised footage itself (for details on narrative approach see Thomas, Lindstedt & Wright, 2005). For the prototype production it was important to identify the classes and range of material a producer would need to aim to assemble when making a reconfigurable programme of this type from scratch.

Once the concept had been developed and the material (3 x 50-minutes) ingested digitally in the BT FMT, descriptive metadata was applied as *Keywords* to specified *Keyword Categories* (Table 1 below) to the 'media objects' (clips) to transform them into 'narrative objects' – the basic building-blocks of new personalisable programmes.

Table 1. Example of fields used to generate the metadata descriptor tags for a media-clip

Metadata Field Descriptor	Metadata Keywords
Section 1	***Helps describe the subject matter and form of the clip***
Topic (what is it about?)	Planets, Stars, The Orient, Time, Gods, Milky Way, Aquarius, Capricorn, Constellations, Gods come to Earth, Water, River
Characters/subjects (who/what?)	Jade Emperor; Chi Nu; Chen Yun
Action (does what?)	The Jade Emperor sends Chi Nu to Earth; she bathes in the Silver River, and the Ox-Herd Chen Yun (the star Altair) steals her garments. The Ox is the constellation Capricorn. Chen Yun carries off Chi Nu and marries her. The Jade Emperor draws the silver River up into the sky to form the Milky Way
Action Type	Pursuit, Theft, Marriage
Place /Location/Culture (where?)	China
Props (with what/how?)	Ox (Capricorn), Silver River (Milky Way); Weaving-Loom (Aquarius)
Genre (Style) (*in what style?*)	Puppet Play
Dialogue (**Voice-Over**)	Voice Over: In Chinese mythology the cosmos is described through the figure of the Jade Emperor, who is absolute ruler of the heavens as the Emperor of China is of the Earth. His daughter, Chi Nu is the celestial Goddess of the Star Vega – her Weaving Loom is the constellation Aquarius. She made the clouds; her garments change colour with the seasons…
Sound	Music, Voice-Over (Harp music over puppet-play and under Voice-Over)
Section 2	***Helps describe the appropriateness of the clip to different modes and complexity levels according to a numerical scale of values***
Engagement Mode	Entertainment 6, Education 5, Information 4
Complexity Level(s)	1, 2

The *Lohse-programme-structure* system was used as a basis for handling *Programmes Templates* in the *GiSC* material (Figure 1 below). Narrative *relationships* and *dependencies* between clips or groups of clips, and *cause/effect* rules are described through a graphical interface at media-clip level. Relationships always take precedence over programme-structural rules, so that critical story and narrative structures will always be maintained (e.g. a point-of-view shot can be marked, so it will always be preceded by – or followed by - a close-up of the appropriate (subject) character's face, Figure 1 below).

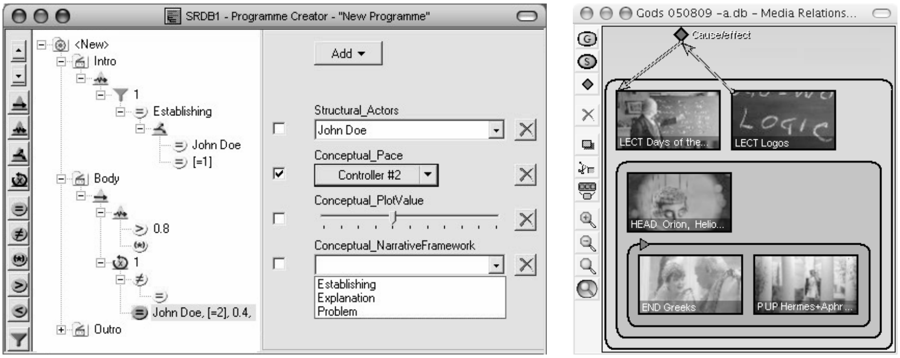


Fig. 1. Implemented *Lohse-programme-structure* system (Ludvig Lohse 2002) and *Relationships Builder* enables the author to define programme and clip specific narrative rules

To enable the creation of the three modes: *Entertainment, Education* and *Information* and offer three Levels of difficulty/complexity within the Educational material, each episodic segment (clip) is described using tags that indicate the subject matter and form of the clip, the suitability of the media clip to a particular mode, level of complexity in *Education Mode*, as well as the best way to ensure good narrative sense when reconfigured (see Table 1 and Figure 1: *Relationships Builder*).

In *GiSC*, the basic argument of programmes follows the documentary rhetorical structure of: Exposition, Development, Summation, Resolution, Conclusion (as in the original programmes); but there is much greater flexibility for associative cross-referencing between topics and themes. The programme formats established in *GiSC* will hopefully form the basis for standardised programme- or format-specific templates which can be saved and applied to new projects.

The script-testing functionality of the upcoming NM2 Toolkit aims to enable the author(s) (writer/director/producer) to test all reconfigurations into plot-frameworks (programme templates) at script-stage, before the material is shot, to ensure that the footage is optimised for this new media content form. The script can be tagged with metadata before or during shooting, so the keywords and programme-templates can be prepared and tested as soon as the script is available. Since scripts tend to follow an established standard, the NM2 system proposes to interpret the script for automatic metadata extraction to expedite effective tagging and improve template-design.

The original *Gods in the Sky* TV-programme was successfully executed through its presentation and content which was targeted at prime-time 'sit-back' livingroom viewing. It is thus inherent that the reconfigurable version compliments the intentions of the original programme (which is similar to many other traditional docu-drama programmes) and creates a similar user-experience through an interface which is easy to use. A hierarchical selection structure offers an uncluttered graphical interface which is controlled via the WMC remote controller using only the five basic navigation buttons (four arrow keys and the 'OK' button, see Figure 2).

The tri-modal programme-format requires different levels of interaction engagement ranging from a single-click and no graphical interface in *Entertainment* mode to an icon-based user-interface in *Education* mode and icon/textual based user-interface in *Information* mode.

Fig. 2. Screen-grab of fully working prototype (*Education mode*) and Windows Media Center remote controller highlighting the five basic navigation buttons (Ludvig Lohse 2005)

3 Conclusions

The conceptual approach to content of determining ‘plot frameworks’, establishing narrative relationships and dependencies, and tagging media-clips with metadata to convert them into ‘narrative objects’ capable of machine-handling was successful in delivering personalised, coherent, satisfying and revisitable programmes using the same pool of media material in three modes (Entertainment, Education and Information). It was established that authoring by programme-makers is necessary in the same way as for conventional documentary, but that a purpose-built toolkit can handle post-production and delivery of the production effectively and efficiently.

GiSC succeeded in creating an interactive livingroom ‘sit-back’ experience using the screen grammar of television, rather than that of a desktop ‘lean-forward’ experience, by employing an uncluttered interface that is easy to use, controlled via a WMC remote controller.

The BT FMT (to be replaced by the more powerful and flexible NM2 Toolkit) proved capable of delivering a fully-working pilot demonstrator which handles 150 minutes of video to deliver personalised versions (‘personal narratives’) from a pool of content (‘global narrative’) with coherent narrative structure.

Acknowledgements

With thanks to WagTV and especially Martin Durkin for permission to use material from 'Gods in the Sky' (broadcast on Channel 4 TV, 2003).

References

1. Thomas, M. and Lohse, L.: Flexible Media: producing character-based reconfigurable drama. In: *Digital Creativity* Vol. 15, No.2. Taylor & Francis: (June 2004) 98-109
2. Thomas, M., Lindstedt, I., and Wright, T.: An introduction to the Narrativity in the interactive environment. EC Report NM2 D2.1 – Public Deliverable. http://www.ist-nm2.org (2005)

Interactivity

The Role of Tangibles in Interactive Storytelling

Ana Paiva

GAIPS Grupo de Agentes Inteligentes e Personagens Sintéticas,
INESC-ID and IST, Tagus Park,
Av. Prof. Cavaco Silva, 2780-990 Porto Salvo, Portugal
ana.paiva@inesc-id.pt

Interactive storytelling is a new area of computing, that brings together computers, interaction and storytelling, with the goal of leading viewers to become active participants in computer supported narrative. Stories become dynamic and emerge in ways perhaps never before explored.

Different approaches can be found on the notion of interactive narrative. According to [1] "the presentation of a narrative can be thought of as movement through a narrative space. In a non-linear narrative there is more than one way to traverse that space. [...] Thus, an author of non-linear narrative must construct a representation that constrains the audience to interesting paths while giving them the freedom to deviate from the unchanging line drawn by traditional linear narrative". As such, when interactivity is brought into storytelling, a quite delicate balance needs to be found between the user's participation, thus interactivity, and the wishes and intended directions of the authors. So, the author's role is not only modified but taken further, into one that allows and provides the interleaving stories, content and constraints for the audience to experience. On the other hand, the audience becomes more than passive viewers, being able to act, and thus making a difference in the way the story progresses. Getting engaged in an interactive storytelling experience is not only moving beyond the simple passive role of experiencing events and situations in a third person perspective, but also taking a more active role in feeling what characters feel and influencing the outcome of the story.

But, to do so, users must be able to interact with the system influencing the outcomes. Several different approaches have be explored for such interaction. For example, Cavazza et. al. [6] use a "God-like" approach, where the user is an invisible avatar in the 3D environment and is able to remove and displace props of narrative significance to change the direction of the story. Machado et. al. [2] proposed that users should also also actors, and their influence in the story is through their basic actions and their interactions. Differently, the approach taken in Façade [5] allows for users to influence the story through natural language conversations established with the main characters, thus influencing its outcome. Another approach is followed in Carmen's Bright IDEAS [4] where the user influences what happens by selecting the thoughts of Carmen, the main character in the story (which appear in form of balloons). Finally, in the case of FearNot! [7] children become friends of a bully victim (the main character in the story) and interact with it by suggesting ways of dealing with his/her problems. In

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 225–228, 2005.

FearNot! the user influences the way the story develops by providing goals and suggesting coping mechanisms to the main character. All these systems, in spite being quite advanced in the way users can influence the interactive narrative, rely mainly on keyboard input.

Another alternative is to allow users to explore all their senses and use their body in their participatory story building. As such, multi-modality can become an emerging technology for engaging participants in computer based interactive storytelling. The physical environment of the participants can then be seen as part of the interface with the virtual world, where stories become a combination of shared worlds' experiences.

Furthermore, objects in the real worlds can also become gateways to the virtual narrative experiences. A participant can use a pen, which, in the virtual world, may turn out to be a fantastic magic wand. The reality of the physical world can be challenged, and somehow augmented with this new form of interaction.

In order to explore the role of tangibles and tangible interfaces for user participation in interactive storytelling, one must challenge the borderline between the real and the virtual. The main question to ask is, how can we build really gateways to the virtual world that do actually work? As we have seen, in most cases the interaction between interactive narrative systems and users presupposes that there is a clear boundary between the characters in the "virtual" story (which live in the virtual world) and the user (that is living in the real world) but at the same time influencing the virtual world.

However, with the introduction of agency and character based approaches, users also "live" in the two worlds [8]. But, despite this dual presence of users, there is still an absence of seamless couplings between these two parallel existences leaving a great divide between the two worlds (see [9]). This division is behind the growing the area of research on "tangible user interfaces" where the goal is to "go beyond the current GUI (Graphic User Interface) and bring current objects to the interaction between users and computers"[9]".

One simple and yet fascinating example where tangibles are used in interactive storytelling is the Lego© product "The story builder" where the physical and the virtual get mixed up, and objects are used in a constructive way to build stories. The Story builder allows children to use plastic cards (the Story Cards), that contain pictures of characters, and by inserting these cards in the system the stories change accordingly. The different cards allows for a set of different possible stories to be made. Although very interesting, the system itself does not really engage children 3 to 5 years old into exploring the different alternative stories. Although the cards are fascinating for the children, their role in the storytelling experience gets somehow confused (one three year old considered the cards as coins, for example).

Other systems have been developed where cards are used as a way to influence the way the story progresses. One of such systems is Papous (see Figure 1) [10] an old granddad that tells stories. Papous can change not only the "mode" in which the story is told (making it more happy or sad, for example) but also the

Fig. 1. Influencing the stories Papous tells through cards

characters and scenes in the story. To make Papous change, children must insert different cards (that picture not only scenes, but also characters) into a kind of mailbox [11].

Children that interacted with Papous felt not only that they influenced the story being told by the character, but particularly, that the physical interaction was immensely enjoyable, and the enthusiasm and curiosity about the way it worked was obvious (see [10] for results).

Another example recently explored was the use of a doll shaped interface (SenToy) that allows users to influence the emotions of a character (avatar) in a 3-D system. SenToy initially worked as an interface to the role playing game (FantasyA) where players would exhibit a particular set of emotions and

Fig. 2. Using SenToy for emotion expression in virtual storytelling

perform a set of actions as a way to evolve in the game (see [8]). The aim of SenToy was to "pull the user into the game" through the use of a physical, touchable affective interface (see Figure 2). With sensors in its limbs, sensitive to movement and acceleration, SenToy captures certain patterns of movements from the users, which are associated with particular emotional expressions. Given the novelty of this interaction, we decided to explore and test if users were able to manipulate SenToy appropriately in an interactive storytelling situation. So, SenToy's components were embedded into the FearNot! storytelling application. Users were allowed to use SenToy and express what they "felt" along the story being shown. The results obtained (see [12]) were indeed very positive, not only because it confirmed once more that children really liked the tangible interface, but also because it allowed us to gather emotional data about the users while interacting with the interactive storytelling environment: FearNot!. In fact, the results show that children did express the emotional states that the authors tried to evoke with the system.

References

1. Galyean T. : Narrative Guidance of Interactivity. MIT Ph.D. Thesis. 1995.
2. Machado, I. and Paiva, A. and Prada, R.: Is the wolf angry or just hungry? Inspecting, modifying and sharing Characters' Minds. In: Proceedings of the International Conference on Autonomous Agents, ACM Press, 2001
3. Sobral, D. and Paiva. A.: "Machiavellian Characters and the Edutainment Paradox", in Intelligent Virtual Agents, Springer, 2003.
4. Marsella, S. and Johnson, W. and LaBore, C.: Interactive Pedagogical Drama. In: *Proceedings of the Fourth International Conference on Autonomous Agents*, ACM Press, 301–308, 2000
5. Mateas, M. and Stern, A.: Natural Language Processing In Façade: Surface-text Processing, *In Technologies for Interactive Digital Storytelling and Entertainment (TIDSE)*, Springer, 2004.
6. Cavazza, M. and Martin, O. and Charles, F. and Mead, S. and Marichal, X.: Interacting with Virtual Agents in Mixed Reality Interactive Storytelling, *Intelligent Virtual Agents (IVA 2003)*, Springer, 2003.
7. Paiva, A. et. al., *Caring for Agents and Agents that Care: Building Empathic Relations with Synthetic Agents*, AAMAS 2004, ACM Press, 2004.
8. Paiva, A. and Andersson, G. and Höök, K. and Mourao, D. and Costa, M. and Martinho, C.: SenToy in FantasyA: Designing an Affective Sympathetic Interface to a Computer Game in *Personal and Ubiquitous Computing* vol. 6. pages 378-389, 2002.
9. Ishii, H. and Ullmer, B., Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms, *Proceedings of Conference on Human Factors in Computing Systems (CHI'97)* ACM Press, 1997.
10. Silva, A. "Papous: The Virtual Storyteller", MSc Thesis, Instituto Superior Técnico, 2004.
11. Sengers et al.: The Enigmatics of Affect, in *Designing Interactive Systems- Conference Proceedings*", ACM Press, 2002.
12. Figueiredo, R. and Paiva, A.: "Watch and Feel: an affective interface in a virtual storytelling environment", to appear in "Affective Computing and Intelligent Interaction", Springer, 2005.

Enabling Communications-Based Interactive Storytelling Through a Tangible Mapping Approach

Norman Lin[1,3], Kenji Mase[1,2,3], Yasuyuki Sumi[1,4], and Tomoji Toriyama[1]

[1] Advanced Telecommunications Research International, Media Information Science Labs, Dep't of Interaction Media. 2-2-2 Hikaridai, Keihanna Science City. Kyoto 619-0288, Japan
[2] Nagoya University, Information Technology Center (Nagoya, Japan)
[3] Nagoya University, Graduate School of Information Science (Nagoya, Japan)
[4] Kyoto University, Graduate School of Informatics, Dep't of Intelligence Science and Technology, (Kyoto, Japan)

Abstract. We present a system supporting the re-use and re-purposing of captured video media for an interactive storytelling performance by a human storyteller. The system's components are based on a philosophy of non-verbal embodied interaction for both media capture and media re-use. The system uses non-verbal interaction sensors during media capture and an augmented reality tangible interface supporting spatial and motion communications affordances during storytelling. We discuss the requirements for an interactive storytelling tool – media retrieval, media re-purposing, and expressive media control – and show how our system design and implementation support these requirements for interactive storytelling. Interface evaluation shows that our system enables communication for storytelling, in particular non-verbal spatial communication. We conclude with a discussion of how our storytelling system is being further developed.

1 Introduction

Historically, new media technologies begin as mass media tools for widespread and impersonal information distribution. As the technology matures and becomes accessible to non-specialists, individuals become able to use the media technology on a personal level. We envision a future where personally controllable ubiquitous and wearable sensors capture large portions of people's daily lives. The availability of such personal media archives offers new possibilities for transmitting and sharing personal experience with others.

We believe that a very effective way to enable experience sharing from such personal media archives is through an interactive storytelling approach. In this paper, we investigate the usage of captured video media for interactive storytelling. Our contributions in this paper are: a general analysis of requirements for communications-based interactive storytelling; presentation of our storytelling system design using interaction sensors and captured video mapped onto tangible artifacts using augmented reality; and user study results indicating suitability of our storytelling interface. Our analysis of storytelling systems differentiates between generative and communications-based approaches; our focus in this paper concentrates on the communications-based approach.

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 229–238, 2005.

2 Related Work

Work on interactive storytelling can broadly be classified into two types of approaches which differ in their focus. One approach, which we call the generative approach, focuses on enabling a computer system to generate stories based on some story formalism. The other approach, which we call the communications approach, focuses on enabling story communication between humans by providing story structuring or performance tools. The generative approach views interactive storytelling as interaction between a human and a computer story system; the communications approach views interactive storytelling as interaction between humans which is mediated through a computer story system.

Cavazza *et al* [1] present a generative approach to interactive storytelling based on the idea of allowing spectator intervention in a well-defined story which is executed by autonomous goal-directed virtual agents. Osborn [2] also adopts a model of user intervention as interaction into a simulated story world with coordinated software agents generating a story as a by-product of agent interactions. Grasbon and Braun [3] present a story engine that provides high-level plot guidance for generating stories based on Vladimir Propp's morphological functions[4, pp. 14-16], derived from the structures of Russian fairy tales.

Examples of the communications approach to interactive storytelling include Balabanovic *et al*'s StoryTrack system[5], which uses stories as way of organizing digital photographs in a hand-held device which can be easily viewed by and passed among different users as a focal point of a conversational storytelling session. Druin *et al*'s PETS (Personal Electronic Teller of Stories) [6] allows children to record their personal stories into stuffed toys, which can later be recalled and played back during storytelling. Cassell and Ryokai's StoryMat [7] records children's stories as they play with stuffed animals on a mat; stories can later be retrieved and projected onto the mat when a child – either alone or in the company of other children – plays with the same animals at the same place on the mat. Bayon *et al* describe their work on KidPad [8], which supports children creating stories through drawing, linking, and zooming elements; the story media can then be interactively navigated with a physical "magic carpet" and bar-coded objects. In all of these communications approaches, human users control the structure and flow of the story, but the computer system and artifacts assist in the interactive and communicative storytelling process.

3 System Design to Support Interactive Storytelling

3.1 Philosophy of Experience

We believe experience, from which personal stories are made, is an interactive, sensory, and social process through which individuals create knowledge about the world. Viewing experience as a process specifically excludes the possibility of directly capturing or directly transferring personal experience to another person; we can capture or transfer media or materials from or relating to the experience, but not the experience itself, since experience is a process that occurs in human minds[9]. We also believe that embodied interactions (Dourish in [10]) which take place in a physical space are an important part of the phenomenon of experience.

3.2 Philosophy of and Requirements for Interactive Storytelling

We view a storytelling session, intended to convey a previous experience, itself as an experiential process. For interactive storytelling, we therefore adopt a communications rather than a generative approach. As Glassner [11, pp. 5-6] observes, there is a difference in structure between casual stories and professionally crafted stories with widespread appeal. We do not aim to make cinematic or novel-like stories from individuals' own experiences, mainly because everyday experience tends to lack dramatic structure [12], especially not the highly structured dramatic forms found in works of literature [4]. Instead of trying to fit complex dramatic structures onto dramatically amorphous personal experiences, we instead want to enable casual communications, interactive story performance, and social construction of shared understandings about individual experience.

Given our communications and experiential philosophy on storytelling, we propose that the following features are needed for an interactive storytelling system which follows a communications approach: (1) Navigation and retrieval of raw media for story creation (2) Re-purposing of media (de-contextualization and re-contextualization) and (3) Non-verbal expressive control. We will say that a system offers support for "interactive storytelling" if it offers interactive support for some or all of the above activities.

Navigation and Retrieval of Raw Media. A storytelling system should provide some support for an indexed or organized way of navigating and retrieving video media so that the storyteller can find useful and interesting story media.

Re-purposing of Media. Re-purposing refers to the de-contextualization of the captured media from its original context, in order to enable re-use of the media during storytelling in new contexts (re-contextualization). We can identify three kinds of re-contextualization: temporal, spatial, and expressive.

Temporal re-contextualization of video refers to the re-sequencing of video footage into another presentation order. Due to the so-called Kuleshov effect (named after Russian film-maker Lev Kuleshov, 1899-1970), the temporal playback order of video clips can greatly influence viewers' perception of the video's meaning.

Spatial re-contextualization of video refers to the separation, within the 2D image space of video frames, of foreground objects' video from the spatially surrounding background scene. In this way, the video of the foreground objects can be used in isolation and placed in a different spatial contexts, allowing re-use of media for many different story purposes.

Expressive re-contextualization of video refers to the association of new, expressive context with an existing segment of video media. This directly relates to the next requirement of an interactive storytelling system.

Non-verbal Expressive Control. Consider a video clip of a person sitting at a desk. Viewed by itself it is neutral. But by superimposing a video effect of a glowing light bulb appearing next to the persons head, we can non-verbally convey the notion that the person suddenly has a "bright idea." By controlling the speed of appearance, size and brightness of the bulb, position, duration of visibility, and so forth, we can further convey nuances of meaning. These additional nuances of meaning are an additional expressiveness that gives new meaning to – and therefore adds new context to – the original video clip. This expressiveness, analogous to the use of emoticons in text, is also a form of re-contextualization. A storytelling system supporting this kind of expressiveness allows for expressive and varied interactive storytelling.

4 System Components for Experience Capture and Interactive Storytelling

Based on the above philosophy of and requirements for interactive storytelling, we now present our system design and system components for interactive storytelling. Our system uses captured video indexed by interaction information and an augmented reality tangible interface designed to allow an interactive storytelling performance using video media in an object-centric way.

4.1 Experience Capture and Media Navigation/Retrieval

As the source of media for our stories, we use an Experience Capture[1] System [13] consisting of wearable and ubiquitous sensors mounted throughout an environment of interest and worn by individual users (Figure 1a). With this system, we can capture video from third-person and first-person viewpoints. Through the use of infrared (IR) trackers – which are head-mounted and environmentally-mounted – and IR identification tags – which are mounted on people and objects – we can detect non-verbal interactions among participants, such as gazing, spatial co-location, or higher-level composite non-verbal interactions [14]. These interaction events are all saved with time-stamps into a database and are used as an index on the captured video data.

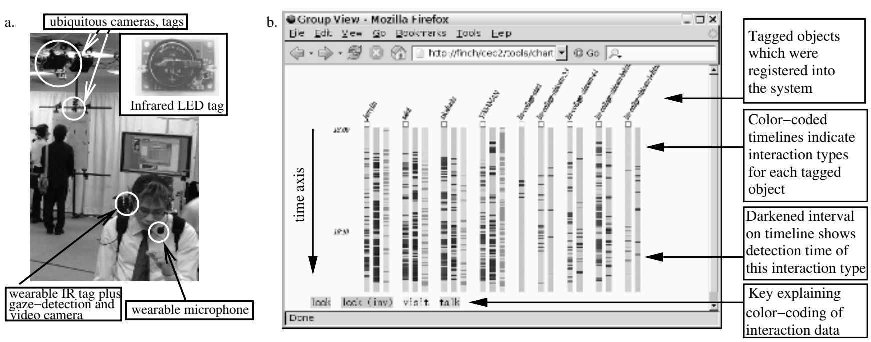


Fig. 1. (a) Experience Capture System. (b) Web-based video browser indexed by interactions.

[1] Our philosophy of "experience-as-a-process" means that the experience itself is not captured, only media and information about the experience. During a later storytelling process, when a storyteller views and manipulates the previously captured media or information, the process creates a new experience in the minds of the story audience.

Media navigation and retrieval are supported through a database browser which allows navigation and retrieval of video segments using the interaction information indexes. The application runs in a web browser using a traditional point-and-click graphical user interface (Figure 1b).

4.2 Re-purposing Tools

We mentioned earlier the need for different kinds of re-purposing captured media for interactive storytelling. Currently we support spatial re-contextualization through an image processing tool, and temporal re-contextualization through a tangible interface affording interactive selection and playback of video clips.

Background Subtraction for Spatial Re-contextualization. To enable spatial re-contextualization, we implemented a tool for background subtraction of foreground video images from a static unmoving background. This can remove, to some extent, unwanted visual clutter and context from the foreground objects, thereby allowing the video of the foreground object to be re-used in new spatial contexts in a storytelling scenario. Figure 2a shows the tool.

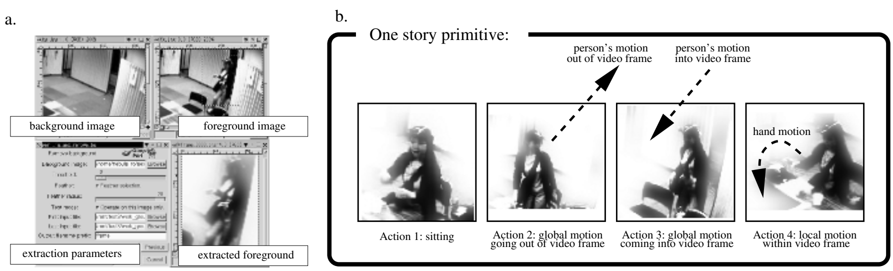


Fig. 2. (a) Image processing tool for cropping and background subtraction of video frames. (b) Example of a story primitive, representing a person, consisting of four different *actions* which are video clips. Labeled arrows show the motion in each video clip and are not part of the video.

Organization of Video Clips into Story Primitives. Video clips, extracted as described in previous section, are intended to be thought of as short *actions* on a *story primitive* (Figure 2b). A story primitive thus conceptually represents one object of interest which can be manipulated interactively by a human storyteller for communications purposes; the organization of video clips is object-centric. Note that this object-centric way of organizing and manipulating video media is a consequence of spatial re-contextualization issues discussed earlier, and is different than purely cinematic story performances using video, such as Lew's Live Cinema [15], which follows traditional cinematic paradigms for video manipulation to enable "assembling [a] film in front of the audience." In contrast, we are not interested in assembling a film, but are rather interested in mapping the video media onto many story primitives which can be manipulated in a real-space to create a new spatial contexts for a storytelling performance. Our approach can be likened to creating video-based puppets from the video media.

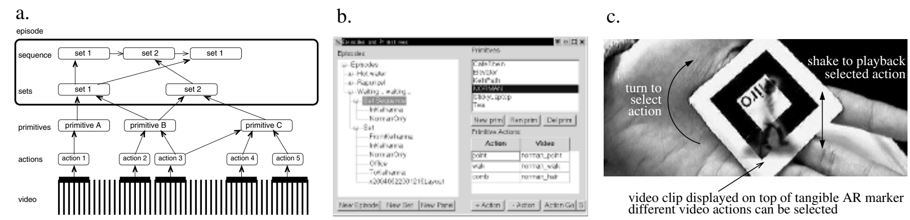


Fig. 3. (a) Organization of story media. (b) Graphical user interface for creating this organization. (c) Mapping story media to an AR tangible interface for interactive manipulation.

After creating a number of useful story primitives, the storyteller chooses a group of story primitives which can be spatially arranged, like props on a theatrical stage, in order to form a meaningful physical stage space in which a story can be told. This group of story primitives is called a *set*, in the sense of a theatrical stage set; several such sets may be created. Sets are grouped into a temporal order called a *sequence*, which indicates the order in which the different sets are to be used during the telling of the story, much like the list of actors and props needed for each act of a play. Together, the sets and the set sequence form a story *episode* (Figures 3a and 3b).

Tangible Storytelling Interface. The last component of our system is the tangible storytelling interface using augmented reality. Story primitives are mapped to physical, tangible artifacts. We use the AR Toolkit [16], a computer-vision-based marker tracking system, combined with an OpenGL-based 3D rendering engine to display the video media from the story primitives on top of the display of the tangible artifact. Viewing the AR display can be done through a head-mounted display or through a conventional computer monitor. Rotating the tangible artifact about its vertical axis, like turning a dial, selects among the different actions assigned to a story primitive (Figure 3c). Shaking the primitive up and down initiates playback of the video for the currently selected action. Interactive, real-time control over the activation and temporal order of video playback allows temporal re-contextualization.

Different story primitives can, through positioning of the tangible artifacts, be physically placed in new spatial locations relative to one another which allows spatial re-contextualization of the video media. Since the tangible artifacts are tracked in real-time, motion of the story primitives can also be communicated. Thus, the tangible storytelling interface offers non-verbal communications affordances which we believe allow effective story communication.

5 System Evaluation and Results

We believe that our storytelling system enables communication which is complementary to spoken storytelling. One important aspect of this complementary communication is non-verbal spatial communication, because of our design choice to map video story media onto tangible artifacts in space. We therefore conducted a non-verbal interface evaluation to test the spatial and motion communications effectiveness of our

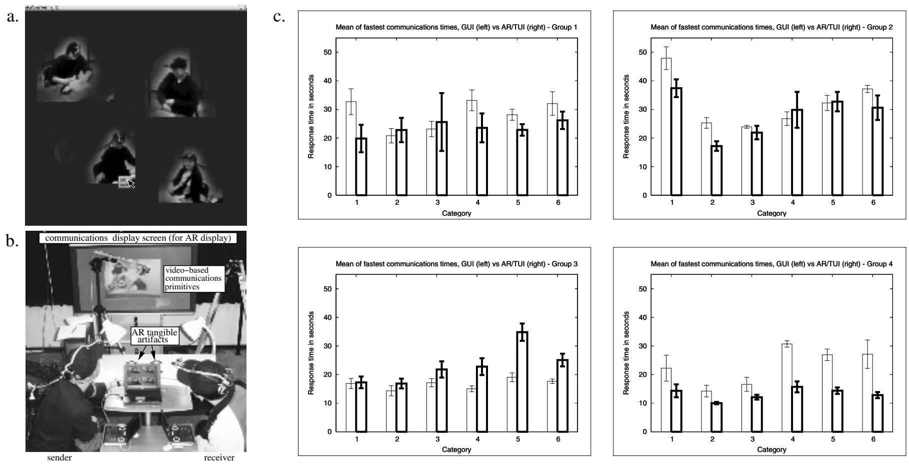


Fig. 4. (a) Graphical user interface for manipulating story primitives. (b) Augmented reality tangible user interface. (c) Mean of fastest communications times for each group. Each group's performance is shown with respect to 6 categories 1-6 (see text for categories). In each graph, the left bar in a category represents communications time required over the GUI and the right bar in a pair represents the communications time required over the AR tangible interface.

augmented reality tangible user interface against that of a traditional point-and-click graphical user interface controlled with mouse and keyboard. We hypothesized that the tangible interface allows more efficient and more accurate non-verbal communication of spatial and motion information than does a graphical user interface.

Our experimental evaluation consisted of two phases. In the first phase, we first recorded video of two groups of four people conducting a structured task in which the persons' locations and motions would change frequently over the course of the experience, which led to many different types of spatial and motion configurations being captured. Figure 2b shows example motions for one person; all four persons in each group, in a pre-defined order, would carry out all of these motions over the course of the experience. There were a total of 34 different configurations for each group of four people, each configuration being distinct from the others in terms of the spatial location and the motion of the participants. After the experience, for each group, video clips of the four individual persons were extracted and aggregated as actions onto four story primitives, which were then mapped onto four tangible artifacts, one for each person, using AR.

In the second phase of the experimental evaluation, four groups of two people (each 2-person group consisting of people from different groups in the first phase) took part in a non-verbal communications quiz. In the quiz, one person was assigned the role of sender; the other, the role of receiver. A quiz program displayed, to the sender only, a video clip of one of the 34 spatial or motion configurations recorded from the original experience of phase 1. The sender's task was to non-verbally communicate the spatial or motion information in this video clip to the receiver. The quiz program displayed, to both the sender and receiver, a multiple-choice answer screen consisting of four different video clips, only one of which was the correct answer (the video clip secretly assigned to the sender). The receiver's task was to interpret the sender's communication and choose the correct answer from the multiple-choice answer screen.

Each multiple-choice answer screen contained 4 carefully selected videos in which the correct answer could be differentiated from the incorrect answers mainly on the basis of a certain kind or *category* of spatial or motion information. A correct answer to a question would imply successful communication of that category of spatial or motion information from the sender to the receiver. In particular, the following six categories of spatial or motion information were tested. (1) Spatial configuration of the group of people: communicate who was present, and where they were located relative to other people. (2) Existence of global motion: communicate that some person walked into or out of the video frame. (3) Agent of global motion: communicate that a specific person walked into or out of the video frame. (4) Direction of global motion: communicate that a specific person walked in a specific direction into or out of the video frame. (5) Existence of local motion: communicate that some person made a movement within the bounds of the video frame. (6) Agent of local motion: communicate that a specific person made a movement within the bounds of the video frame.

We compared communications accuracy and communications time using both a point-and-click graphical user interface and the tangible AR interface to story media. For categories 1, 2, and 3, the means of the fastest communication times over the tangible interface were equal to or faster than those of the graphical interface for all groups; for categories 4, 5, and 6, the same results were visible in 3 of the 4 groups. The mean communications accuracy was comparable over both interfaces. This shows that the spatial communications affordances of the tangible user interface are effective at communicating non-verbal spatial information. We believe that our system, by supporting this kind of non-verbal spatial communication with story media, can and will complement spoken storytelling.

6 Future Work and Conclusion

We are continuing to develop our system in many ways. One issue is that the system does not currently provide support for non-verbal expressive control of media; we can select or playback video clips assigned to the story primitives, but cannot control or enhance the media once it has started playback. As a first step to enable this kind of control, we have implemented an augmented expressive pointer which is designed to allow for real-time, analog, and expressive control of the video playback speed and direction; we also envision that real-time expressive video effects will also be controlled in this manner. The device consists of an AR Toolkit marker mounted on a wireless mouse with a scroll wheel. Physical proximity, in real space, of the augmented pointer to a tangible artifact (marker) causes selection of that artifact and the associated story primitive. Scrolling the mouse wheel forward or backward allows fine-grained stepping-through or "scrubbing" of the video for the currently selected video action. For instance, wiggling the wheel forward and backward allows real-time emphasis or repetition of certain small sub-parts of the currently selected video action in a manner similar to that of a video jockey artist (Figure 5). We are also investigating ways of using the the a.

b.

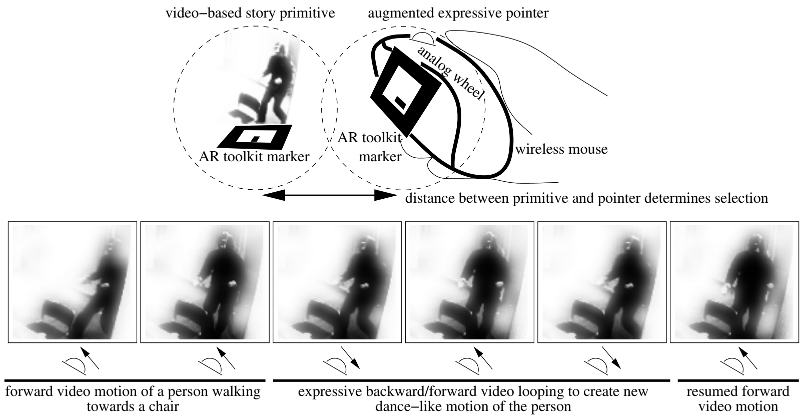


Fig. 5. (a) Augmented expressive pointer. (b) Example of using the analog wheel for real-time expressive video looping of a walking figure.

tangible AR markers alone, without separate hardware such as a mouse, for such analog expressive video control.

In conclusion, we presented an interactive storytelling system based on the idea of storytelling-as-experience and experience-as-a-process. We choose to follow a communications approach (as opposed to a generative approach) in which a non-verbal video-based storytelling system can be used by a storyteller parallel with and complementary to spoken storytelling. Tools enable spatial and temporal media re-contextualization of video media captured from personal experience, and mapping video onto tangible artifacts using AR. Results from a non-verbal user interface study indicate that our system allows effective non-verbal communication of spatial information, which we believe will complement spoken storytelling. Future experiments will further investigate this belief.

We would like to sincerely thank the members of the ATR Media Information Science labs, department of Interaction Media, for their development and support of the Experience Capture System and for valuable discussion. This work was sponsored in part by the National Institute of Information and Communications Technology.

References

1. Cavazza, M., Charles, F., Mead, S.J.: Emergent situations in interactive storytelling. In: Proceedings of 2002 ACM Symposium on Applied Computing (SAC2002). (2002)
2. Osborn, B.A.: An Agent-Based Architecture For Generating Interactive Stories. PhD thesis, Naval Postgraduate School, Monterey, California (2002)
3. Grasbon, D., Braun, N.: A morphological approach to interactive storytelling. In Fleischmann, M., Strauss, W., eds.: Proceedings: cast01 // living in mixed realities, special issue of netzspannung.org/journal, The magazine for media production and inter-media research. (2001)
4. Brooks, P.: Reading for the Plot. Harvard University Press (1998) Fifth printing. Originally published: New York: Knopf, 1984.

5. Balabanovic, M., Chu, L.L., Wolff, G.J.: Storytelling with digital photographs. In: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY USA, ACM Press (2000) 564 – 571
6. Druin, A., Montemayor, J., Hendler, J., McAlister, B., Boltman, A., Fiterman, E., Plaisant, A., Kruskal, A., Olsen, H., Revett, I., Schwenn, T.P., Sumida, L., Wagner, R.: Designing PETS: A personal electronic teller of stories. In: Proceedings of CHI 1999. (1999) 326–329
7. Cassell, J., Ryokai, K.: Making space for voice: Technologies to support children's fantasy and storytelling. Personal and Ubiquitous Computing **5** (2001) 169–190
8. Bayon, V., Wilson, J.R., Stanton, D., Boltman, A.: Mixed reality storytelling environments. Virtual Reality **7** (2003) 54–63
9. Davis, M.: Theoretical foundations for experiential systems design. In: Proceedings of the 2003 ACM SIGMM workshop on Experiential telepresence. Berkeley, California. (2003) 45–52
10. Dourish, P.: Where the Action Is - The Foundations of Embodied Interaction. Massachusetts Institute of Technology (2001)
11. Glassner, A.: Interactive Storytelling - Techniques for 21st Century Fiction. A K Peters, Natick, Massachusetts (2004)
12. Appan, P., Sundaram, H., Birchfield, D.: Communicating everyday experiences. In: Proceedings of the 1st ACM workshop on Story representation, mechanism and context. New York, NY, USA, ACM Press (2004) 17–24
13. Sumi, Y.: Collaborative capturing of interactions by wearable/ubiquitous sensors. The 2nd CREST Workshop on Advanced Computing and Communicating Techniques for Wearable Information Playing, Panel "Killer Applications to Implement Wearable Information Playing Stations Used in Daily Life", Nara, Japan (2003)
14. Takahashi, M., Ito, S., Sumi, Y., Tsuchikawa, M., Kogure, K., Mase, K., Nishida, T.: A layered interpretation of human interaction captured by ubiquitous sensors. Presented at CARPE 2004, The First ACM Workshop on Continuous Archival and Retrieval of Personal Experiences, in conjunction with ACM Multimedia 2004 (2004)
15. Lew, M.: Live cinema: Designing an instrument for cinema editing as a live performance. In: Proceedings of the 2004 Conference on New Interfaces for Musical Expression (NIME04). Hamamatsu, Japan. (2004)
16. Kato, H., Billinghurst, M.: Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In: Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99), San Francisco, USA (1999)

A Multidimensional Scale Model to Measure the Interactivity of Virtual Storytelling*

EunKyung Chung and Elizabeth Figa

School of Library and Information Science,
Texas Center for Digital Knowledge,
University of North Texas,
P.O. Box 311068, Denton, TX 76203
{echung, efiga}@lis.admin.unt.edu
http://courses.unt.edu/efiga/Figa/index.htm

Abstract. A measure of interactivity not only can quantify the construct of virtual storytelling systems and thus make it possible to find the precise relationship between interactivity and other variables, but also can help ensure the effective development process of virtual storytelling systems. This study proposes a multidimensional scale model to measure the interactivity of conceptually and technologically diverse virtual storytelling systems. Using a survey research method, users' perceptions of interactivity of virtual storytelling systems were utilized for validating a scale model. This study suggests that multidimensionality should be incorporated into a scale model such as control, synchronicity, and communication. In particular, the findings of this study showed a possibility of sub-dimensions in the *control dimension* which is mechanical-oriented control and content-oriented control. The sub-dimensions can be explained as diverse features provided to the control dimension of interactivity.

1 Introduction

Storytelling, as an information cycle, features its own particular and unique characteristics and levels of interactivity. Storytellers interact with audience in a responsive feedback loop of communication and have the opportunity to morph and change their stories, or the ways of narration and presentation, based on the reactions from story listeners. The digitization or virtualization of storytelling is not merely scanning images, developing multimedia, or converting the contents of information into bunch of '0's and '1's. Rather it is incorporating the characteristics of *interactivity* into a storytelling system. Accordingly, various types of technologies have emerged to introduce and/or expand interactivity in virtual storytelling systems. This study addresses research questions concerning the quantitative identification of the interactivity between audiences and storytelling systems. Researchers and developers likely assume that it is favorable for users/learners to immerse into a storytelling system. This provides more engagement with the system and simulates the real-time multidimensional and

* We are grateful to Texas Center for Digital Knowledge and School of Library and Information Science at the University of North Texas for funding this study and to the anonymous reviewers for their helpful comments on the earlier versions of this paper.

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 239 – 248, 2005.

interactive storytelling experience. In the case of improving comprehension of the story content, even the simplest type of interactivity is more welcomed [15]. However, there has remained uncertainty about how much more or how much less inactivity to incorporate. Some research questions worthy of consideration in examining levels of interactivity of a digital or virtual storytelling systems include: What level of interactivity is appropriate for a particular group of target users? How much interactivity would be appropriate for a particular type or genre of story? What degree of interactivity does an audience expect from a specific virtual storytelling system? We believe all of these are legitimate questions for virtual storytelling researchers and designers and that they can be answered with a reliable and valid scale to measure levels of interactivity. With appropriate measurement tools, analyzing the interactivity of virtual storytelling systems can be accomplished in a reliable and consistent manner. The goal of this study is to develop a scale model to measure the interactivity for virtual storytelling systems and demonstrate its validity and reliability.

2 Interactivity in Plot, Narrations, and Question/Answering

In a sense, incorporating interactive characteristics into virtual storytelling systems is achieving the virtualization of storytelling. In view of that, technologies have been developed in various forms: digital recordings of live storytelling performance; virtual reality-based, augmented reality-based and artificial intelligence-based conversational agents; and interactive narrative hypermedia. Storytelling systems with diverse technological designs provide users with different levels of interactivity. While some allow users to engage in deep immersion with a storytelling system (such as taking on a role in the story), others permit users to operate a storytelling system with the level of VCR- or computer game-type interactivity. Planning for and incorporating interactivity as goal in their system designs, virtual storytelling developers and researchers have manipulated the construct in various ways. However, the research seems to have a gap in identifying specific measures for interactivity. The lack of consistent interactivity level identification can result in misleading assumptions by virtual storytelling developers and researchers. Quantified measurements of interactivity may reveal the need for changes and open up opportunities to aid in achieving the goals or objectives the interactivity level of virtual storytelling systems. For instance, quantified measurements of interactivity can reveal the level of interactivity that actually facilitates user involvement with a particular system or design feature.

In various domains, the definition of interactivity is dissimilar as to what interactivity actually means and represents. However, with respect to human-computer environments, interactivity can be regarded as the degree to which users of a medium can influence the form or content of the mediated environment. In this sense, researchers have pursued different types of approaches to achieve the interactivity of storytelling tools, games, and environments. While some have focused primarily on the interactivity of the plot level, others focused on the interactivity of narration and presentation styles. Other approaches involve interactivity for question-answering using knowledge bases such as WordNet and OpenMind.

The first category of research under consideration concerns providing users with interactivity of plot modification. While the common storytelling types are linear timeline based stories, researchers with this type of interests focus on story modification based on hierarchy and apply a multi-dimensional graph structure [2] to the system. The structure offers branching at certain points in the narrative, to allow users to morph their own plots within the system. The approach to the interactivity of plot modification according to users' interactions could be considered separately as script-based, character-based, and hybrid type.

Secondly, some researchers consider the narration and presentation style aspects as the main target to achieve the interactivity for virtual storytelling. In natural communication settings, narration and presentation styles have the potential to transform the content of a simple story into a fascinating narrative to engage and immerse a participant into a story. One such approach is based on embodied conversational agents which possess the possibilities of becoming a synthetic storyteller [3], [12] and others provide a demonstration with more concentrated interactivity for a virtual storyteller [14]. Users' input influences or acts upon the characters behaviors such as their facial expressions, voice, and gestures. For instance, users can manipulate a story to be narrated in a more scary way resulting in an embodied character's ability to express himself with a visibly scary behavior or appearance. There are some considerations of the interactivity inhabited in virtual environments [17]. In contrast with other virtual environment approaches, they take the narrative and presentation levels into account in order to achieve interactivity. Another fundamental interactivity feature in narration and presentation is manipulating the viewpoints. These systems focus on fulfilling and responding to user-specified viewing requests [1]. Finally, there have been approaches focusing on the "aboutness" of stories that have provided a solution for answering questions related with the content of stories [6]. The answering of questions from users can be considered one of the most explicit verbal-based interactivity levels for storytelling.

Researchers have focused on achieving interactivity through a variety of aspects such as plot, narration, presentation, and question-answering with users. While virtual storytelling systems achieve the interactivity through their design and features, users perceive interactivity as one whole entity rather than as only one specific aspect, for instance, interactivity of narration style. There is an approach to analyze interactivity in terms of *virtual reality* that explores previous research related to interactivity [13]. This review, however, only explores the *qualitative aspects* of interactivity from the existing virtual storytelling systems in the learning processes, and does not provide any quantitative measurement tools for comparing different storytelling approaches. Some quantitative approaches to measure interactivity have been devoted to the study of the Website interactivity. Those studies have focused on identifying scales to measure the interactivity of Websites and their effectiveness for users [19], [4], [11].

3 Dimensions of Interactivity and Research Design

3.1 Dimensions of Interactivity

Primarily, interactivity has been considered multidimensional. Several studies have identified a range of dimensions of interactivity. Generally, interactivity has been approached from the perspectives of communication or cognition. As shown in Table 1, various dimensions identified from a variety of disciplines can be categorized by Liu

[10]'s three major dimensions, *control*, *communication*, and *synchronicity*. Liu's dimensions are selected because they have tested for the interactivity on the Web environment and are related to this study. First, the Control Dimension is considered as Playfulness [7], Choice [7], Complexity of Choice Available [8], Effort to Access [8], Control [18], Exchange of Roles [18], Range [16], Mapping [16], and Active Control [10]. The concept of Control Dimension is comparable to Significance [9], and it impacts user perception of how much influence (or power) he has over the system depending on the options available. Secondly, the dimensions of Communication were identified based on the types of communication [11] and these were recognized as Monologue Communication Dimension, Feedback Communication Dimension, Responsive Dialogue, and Mutual Discourse. These primarily include interaction between a user and a system and facilitate the users' feedback into the system. As to the monologue communication aspect, Information Collection [7] and Monitor System Use [8] are identified. As to feedback communication aspect, Easy of Adding Information [8] is identified. Reciprocal Communication [7], Interpersonal Communication [8], Mutual Discourse [18], and Two-Way Communication [10] can be regarded as mutual discourse communication aspect. Thirdly, as the Synchronicity Dimension, Connectedness [7], Use Responsiveness [8], Speed [16], and Synchronicity [10] are identified. Synchronicity represents the interaction speed between a user and a system. For instance, users measure the interactivity according to how promptly a system responds.

Table 1. Interactivity Dimension

Dimension	Identified Dimension
Control	Playfulness [7], Choice [7], Complexity of Choice Available [8], Effort to Access [8], Control [18], Exchange of Roles[18], Range [16], Mapping [16], Active Control [10]
Communication	Information Collection [7], Reciprocal Communication [7], Monitor System Use [8], Easy of Adding Information [8], Interpersonal Communication [8], Mutual Discourse [18], Two-Way Communication [10]
Synchronicity	Connectedness [7], Use Responsiveness [8], Speed [16], Synchronicity [10]

3.2 Scale Items in Dimensions

Based on three dimensions by Liu [10] such as Control, Communication, and Synchronicity, scale items are selected. In addition, the scale items reflect the features of virtual storytelling systems provided from various technological designs. As shown in **Table 2**, the three dimensions are described with their constituent scale items. While the items in Liu's study [10] are selected from features on the Web, the constitute items in three dimensions are selected from features of virtual storytelling systems as cited in the literature ([1, 3, 12, 14, and 17]). The Control Dimension consists of subdimensions such as Overall Control, Narration Style or Presentation Control, and Plot

Control. While Overall Control addresses the interactivity of perceived feelings of control over the system, Narration Style or Presentation Control describes the interactivity of changing the style of a story, for instance, changing a narration style in a gentle way or a dramatic way. Plot Control concentrates on the interactivity of choosing or directing a different plot line at a certain point in the story. The Communication Dimension discerns different types of communication such as Monologue, Feedback, Responsive Dialogue, and Mutual Discourse [11]. While Monologue addresses the interactivity of obtaining responses from the system, Feedback is related to inputting users' thoughts into the system. In addition, while Responsive Dialogue is described as the interactivity of asking questions and obtaining answers about stories, Mutual Discourse is related to the interactivity of asking questions and obtaining answers during a storytelling experience. Mutual Discourse is more focused on providing users with the initiatives to start the communication holding concurrent and shared understandings. The Synchronicity Dimension primarily explains the interactivity related to the system response times to a user such as animating virtual actors in real-time, special effecting in real-time, and processing users' actions in real-time.

Table 2. Dimensions of Interactivity and Constitute Items

Dimension		Item
Control	Overall Control	Overall Control Feeling over a System Stop/Resume Control Moving the story lines back and forth Changing viewpoints Following a specific virtual actor
	Narration Style or Presentation Control	Changing a narration style (gentle or scary)
	Plot Control	Changing a plot at a certain point
Synchronicity	Real-time responses	Animating virtual actors in real-time Special effects in real-time Processing in users' actions in real-time
Communication	Monologue	Obtaining Information from a System
	Feedback	Inputting a User's feedback to a System
	Responsive Dialogue	Questioning and Answering
	Mutual Discourse	Questioning and Answering during storytelling

3.3 Research Design

This study conducted exploratory research on the development of a multidimensional scale model to measure the interactivity of virtual storytelling systems. A survey instrument method was administrated to evaluate human subjects' perceptions of interactivity for virtual storytelling systems. The participants were randomly recruited from library users at the University of North Texas during the period of May 9 through May 14, 2005. Once pre-tested, the survey tool (available on the Web at http://courses.unt.edu/efiga/ChungFigaResearchQuestionnaire_Approved.htm) was administered to 86 participants from which 83 sets of responses were usable for analysis. Selected scale items in dimensions based on interactive storytelling system features were reviewed by the participants. Each participant was then asked to evaluate the scale items in three dimensions as a numeric value from 1 to 5 with 5 being the highest rating for interactivity. For instance, if a user perceives a scale item in the Control Dimension as a very strong item for interactivity, the user was asked to respond the scale item with the highest point rating, which is 5. The results of the survey were used for data analysis. For a reliable scale, the data were examined for the relationship of scale items and each dimension and for the overall reliability. Consequently, the data analysis consists of two main parts. First, parametric correlation was used to test the relationship between items and each dimension. In addition, in order to show that how much influential each item has, factor analysis was used. The factor analysis is usually used to validate a scale or index by demonstrating that its constituent items load on the same factor. Finally, we used a reliability analysis model, Cronbach's alpha [5], to determine the consistency of three dimensions of interactivity based on the inter-item correlation. Cronbach's alpha is selected because it is generally used to demonstrate the reliability or consistency of the score variations given by participants for each dimension.

4 Results

The 83 responses were analyzed for developing a multidimensional scale model to measure the interactivity of virtual storytelling systems. We surveyed randomly selected non-experts in storytelling because they are representative of a general population of storytelling interactors. **Table 3** shows the demographic information of the study participants. The participants are primarily a younger generation with technology acumen and regularity of computer access and use. Approximately 90% of participants are between 18 and 35 years old. 74% of participants have used computers over six years. Currently, 84% of participants use computer between 6 and 40 hours per week. However, more than half of participants (61%) do not have experience attending or participating in any type of storytelling event.

4.1 Relationship Between Scale Items and Dimensions

The numeric values assigned by participants according to their perception on interactivity were analyzed using Factor Analysis of the SPSS package. Factor analysis is generally used to determine whether there is a relationship between variables (items) and associated dimension or other dimension. As shown in **Table 4**, item-to-associated dimension correlation and inter-item correlation within each dimension were calculated. First, the item-total was computed to illustrate the relationships between scale items and dimensions. Below .50 in item-total are considered to be weak in the relationship between the item and the dimension. There is one problematic item in the Control and Synchronicity Dimension, respectively. The item below .50 in the Control Dimension is item 5, following a specific virtual character during storytelling. Since the activity of following a specific character is involved, it assumes that users don't perceive the item as within their control over the system. Based on the

Table 3. Demographic Information of the Study Participants

Item		Percentage
Gender	Female	62.3
	Male	37.7
Age	8-25 years old	66.7
	26-35 years old	23.2
	36-45 years old	5.8
	over 45 years old	4.3
Years using computer		
	Under 2 years	1.4
	2-4 years	4.3
	5-6 years	20.3
	over 6 years	73.9
Time spent on computer per week		
	Under 5 hours	16.0
	6-20 hours	44.9
	21-40 hours	26.1
	over 40 hours	13.0
Off-line storytelling experiences		
	Never	60.9
	Under 5 times	31.9
	6-10 times	2.9
	11-20 times	1.4
	Over 20 times	2.9

correlation with other items, it was more related to the Communication Dimension. In the Synchronicity Dimension, the item below .50 is processing users' actions in real-time. Based on the correlation of items, the item was more related with the Control Dimension, instead of the Synchronicity. Supposedly, users perceive the processing of users' actions as their control feelings over the system. Secondly, Pearson correlations between all items were calculated to see if there are the relationships between items and unassociated dimensions. Two pairs of items were identified as the relationships with unassociated dimensions. As item 5 in the Control Dimension showed a higher relationship with the Communication Dimension, item 10 showed a stronger relationship with the Control Dimension than associated dimension. Consequently, item 5 and item 10 in each dimension was shown to be problematic and more related with unassociated dimensions. Thirdly, items within each dimension were factor analyzed. In addition, **Table 4** shows factor loadings in each dimension. As the relationship between items and dimensions illustrated, item 5 in the Control Dimension and item 10 in the Synchronicity Dimension are less influential than other items.

More importantly, the results of factor analysis in each dimension reveal sub-dimensions under the control dimension. While there was an expectation of separate sub-dimensions such as overall control, narration style or presentation control, and plot control, the participants perceived the Control Dimension as mechanical control and content-oriented control. According to the Rotated Component Matrix performed

Table 4. Inter-to-Total Correlation and Factor Analysis of Items within Each Dimension

Dimension	Item	Item-to-total	Factor Loading
Control	1. Overall Control Feeling over a System	.505	.540
	2. Stop/Resume	.578	.677
	3. Moving the story lines back and forth	.589	.779
	4. Changing viewpoints	.502	.517
	5. Following a specific virtual actor	.402	.395
	6. Changing a narration style (gentle or scary)	.555	.593
	7. Changing a plot at a certain point	.653	.767
Synchronicity	8. Animating virtual actors in real-time	.787	.737
	9. Special effects in real-time	.645	.586
	10. Processing in users' actions in real-time	.455	.441
Communication	11. Obtaining Information from a System	.644	.662
	12. Inputting a User's feedback to a System	.621	.645
	13. Questioning and Answering	.576	.593
	14. Questioning and Answering during storytelling	.528	.534

using the SPSS Factor Analysis of item 1 through 7 (Table), it revealed two sub-dimensions inside the dimension: item 1-3 and item 4-7. Items 1 through 3 characterize mechanical-oriented control over a system. They contain items such as "overall control feeling", "stop and resume at any time", and "moving the story back-and-forth freely". The other sub-dimension in the Control Dimension is related to the content-oriented control of presentation, viewing, narration style changes, and plot changes at a certain point.

Table 5. Rotated Component Matrix in item 1 – 7 in Control Dimension

	Component	
	1	2
Item 1: Overall Control Feeling over a System	.051	.744
Item 2: Stop/Resume	.131	.820
Item 3: Moving the story lines back and forth	.248	.791
Item 4: Changing viewpoints	.644	.341
Item 5: Following a specific virtual actor	.698	.045
Item 6: Changing a narration style	.779	.105
Item 7: Changing a plot at a certain point	.838	.143

4.2 Scale Reliability

In order to determine how well a set of items characterize the consistency or reliability in each dimension, Cronbach's alpha was calculated as shown in **Table 6**. The Cronbach's Alpha values are consistent with the results of relationships between

Table 6. The Cronbach's Alpha for Each Dimension

Dimension	Cronbach's Alpha
Control	.626
Synchronicity	.720
Two-Way Communication	.763

items and dimensions. As the Control Dimension showed a possibility of sub-dimensions within the dimension, the Cronbach's Alpha value of the Control Dimension was relatively lower than other dimensions.

5 Conclusion

The consensus among virtual storytelling researchers is that integrating interactivity into virtual storytelling systems is one way of bringing the traditional heritage of storytelling into a myriad of environments for education and entertainment users while retaining many of storytelling's unique characteristics. However, there have been few quantitative approaches to specifically measure the interactivity between users and systems via user studies. By finding methods to more precisely and consistently measure levels of interactivity as perceived by users, researchers and designers can develop interactivity models for virtual storytelling systems more suitable for target audiences. With an intention to develop a scale model for measuring the interactivity of virtual storytelling systems, this study incorporates multidimensionality of interactivity into a scale model: *control*, *synchronicity*, and *communication*. This study demonstrates the validity of the multidimensional scale model using a survey research method. Several items were perceived by the study participants as having different dimensions, respectively. Specifically, respondents perceived the item about following a specific virtual actor as the Communication Dimension instead of the Control Dimension. As well, they perceived the item about processing users' actions in real time as the Control Dimension instead of the Synchronicity Dimension. In particular, the data analysis of this study showed a possibility of sub-dimensions, mechanic-oriented and content-oriented sub-dimension, within the Control Dimension. A possible explanation for this result could be found in various types of options or controls provided as features of virtual storytelling system.

For future research, we propose a study using experts in storytelling to evaluate and rate to the level of interactivity for the 14 items within each of the three dimensions to develop an overall interactivity metric. Lastly, the post-storytelling response cycle as a form of interactivity can be considered for future study.

References

1. Bares, W. & Lester, J.: Intelligent multi-shot visualization interfaces for dynamics 3D worlds. Proceedings of the 1999 International Conference on Intelligent User Interfaces (1999)
2. Bimber, O., Encarnacao, M., & Schmalstieg, D.: The virtual showcase as a new platform for sugmented reality digital storytelling. Proceedings of the Workshop on Virtual Environments (2003)

3. Cassell, J.: Conversation as a system framework: designing embodied conversational agents. In J. Cassell et al. (eds.) Embodied Conversational Agents. MIT Press (2003)
4. Cho, C., Leckenby, J.: Interactivity as a measure of advertising effectiveness. Proceedings of the American Academy of Advertising (1999)
5. Cronbach, L.J.: Coefficient Alpha and the Internal Structures of Tests. Psychometrika 16 (1951)
6. Figa, E., Tarau, P.: Story traces and projections: exploring the patterns of storytelling. Proceedings of Technologies for Interactive Digital Storytelling and Entertainment (2003)
7. Ha, L., James, E.L.: Interactivity Reexamined: A Baseline Analysis of Early Business Web Sites, Journal of Broadcasting & Electronic Media, 42(4), (1998)
8. Heeter, D.: Implications of New Interactive Technologies for Conceptualizing Communication, Salvaggio, J.L., Bryant, J. (Ed) Media Use in the Information Age: Emerging Patterns of Adoption and Consumer Use. Hillsdale, NJ. Lawrence Erlbaum Associates (1989)
9. Laurel, B.: Computer as Theatre. Reading, Massachusetts: Addison-Wesley Publishing Company (1993)
10. Liu, Y.: Developing a scale to measure the interactivity of web sites, Journal of Advertising Research, 43 (2003) 207-216
11. McMillan, S.: Interactivity is in the eye of the beholder: function, perception, involvement, and attitude toward the web site. Proceedings of the 2000 Conference of the American Academy of Advertising, (2000) 71-78
12. Noma, T., Zhao, L., & Badler, N.: Design of a virtual human presenter. IEEE Computer Graphics and Applications, 20(4), (2000)
13. Roussou, M.: Learning by doing and learning through play: an explanation of interactivity in virtual environments for children. ACM Computers in Entertainment, 2(1), (2004)
14. Silva, A., Raimundo, G., Paiva, A.: Tell me that bit again … bringing interactivity to a virtual storyteller. Proceedings of the 2nd International Conference on Virtual Conference on Virtual Storytelling (2003)
15. Steiner, K., Tomkins, J.: Narrative event adaptation in virtual environments. Proceedings of 2004 ACM Conference on Intelligent User Interface (2004)
16. Steuer, J.: Defining Virtual Reality: Dimensions Determining Telepresence. Journal of Communication, 42(4), (1992)
17. Theune, M. Faas, S., Nijholt, A., Heylan, D.: The virtual storyteller. SIGGROUP Bulletin, 23(2), (2002)
18. Williams, F., Rice, R.E., Rogers, E.M.: Research Methods and the New Media. New York, NY. The Free Press (1988)
19. Wu, G.: Perceived interactivity and attitude toward website. The Annual Conference of American Academy of Advertising (1999)

Applications

The Rapunsel Project

Ken Perlin[1], Mary Flanagan[2], and Andrea Hollingshead[3]

[1] New York University, USA
[2] Hunter College, USA
[3] University of South California, USA
`http://mrl.nyu.edu/~perlin/`

Abstract. By use of a dance game, and after much input and advice from thirteen year old design consultants, we teach Java programming in a way that will interest middle school girls - a critical age group for addressing gender inequity in programming.

1 Introduction

The RAPUNSEL (Real-time Applied Programming for Underrepresented Students' Early Learning) project is now entering the third year of a three year grant from the National Science Foundation. This paper serves as a progress report, at a point where much of the capability has been implemented at the Media Research Laboratory at New York University and the Tiltfactor Laboratory at Hunter College, but formal testing (scheduled for the third and final year) has not yet been done. Because of the importance of the problem that the project addresses, and the time-critical nature of research in this area, it was felt that it is important to present a snapshot of the project's progress, even before the results of formal testing are in.

The problem is as follows: There is a critical shortage of women in Computer Science (CS) careers and degree programs. Margolis and Fisher, in their landmark study of gender and CS, note that the male dominance in information technology can be linked to the social, cultural, and educational influences and patterns formed in childhood [2, 17]. Research shows that although girls are as talented as boys in math and science, and although most girls are excited about science in childhood, these same girls begin to lose interest in math and science in middle school [8, 9]. By the eighth grade, twice as many boys as girls show an interest in science, engineering, and mathematics careers [10]. While opportunities for CS-related careers are broadening, and programming skills are required in many diverse fields, fewer and fewer girls are attracted to CS related activities. New resources are needed to engage girls in computer programming activities.

For example, according to an NSF survey, 93% of programmers in the USA are men, and only 7% of programmers are women. This is an enormous gender imbalance. Some studies show the gender imbalance in computer science as the worst among the engineering professions. Why is this the case? To summarize briefly the results of a number of studies: i) There are social and cultural preconceptions about

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 251 – 259, 2005.

science that influence girls' involvement in programming and, later, computer science, ii) Programming is taught outside a larger context, and the reason to learn programming is thus unclear, and iii) programming is taught as a "loner activity". These factors lead to under-involvement from girls among a variety of ethnic and socioeconomic backgrounds.

At the same time the demand for participants in CS is increasing, fewer girls than boys enroll in CS classes, feel self-confident with computers, and use computers outside the classroom-- fewer than 33% of participants in computer courses and related programs are girls [1, 10]. Research shows that females experience a reduction in self-esteem during adolescence, negatively affecting their achievements and narrowing their aspirations [21, 24]. Middle school girls statistically drop out of math and science classes or do not perform well, and adolescence is often the final time girls consider the diverse array of career opportunities in technical areas, especially CS [22]. The most significant reasons cited in research for this lack of interest are that girls often underestimate their own abilities and are not engaged by the content of such programs [4, 5, 14]. Women and girls consistently report a lack of confidence in their computer skills [7, 13, 15, 18, 20, 25]. Although women constitute roughly half of the US population, they are significantly underrepresented in CS degree programs and professions [12]. Nearly 75% of future jobs will require computer use, and yet fewer than 33% of participants in computer courses and related activities are girls; by 2010, the largest industries and fastest growing job opportunities will be computer related, specifically, in CS and engineering fields [23]. The current science, engineering and technology workforce is only 19% female [10].

These figures point to an emergency situation in computer literacy for girls. Girls need to be reached before adolescence to keep them interested in science and math and to foster their achievement in these curricular areas. Specifically, girls need to be encouraged not only to be computer users, but also become those at the forefront of creating new computer technologies. Bruckman et. al. found that gender does not affect programming ability or performance, but while girls spend significantly more time than boys communicating with others in computer supported learning environments, boys are more likely than girls to have prior programming experience, and spend more time programming on average [6]. This research recommends that in order to increase gender equity in technical computer skills, developers should focus on strategies for fostering interest among girls. Yet no commercial software exists to encourage middle school girls to learn software design and computer programming, and the types of software directed at girls has neither increased nor diversified since the seminal work of Laurel [16].

Computer games that currently attract girls are thus important to our research. Games for girls have been studied through commissions sponsored by groups such as the AAUW. A recent review of popular mathematics programs targeted at K - 6th grade showed that only 12 % of the characters were female, and those played passive roles such as "princess" [3]. Studies of gender and play have shown that girls are more likely than boys to engage in parallel and constructive play as well as peer conversations [19]. Game environments such as The SIMS and NeoPets engage players in interesting variations of constructive play - players create virtual households or create and care for virtual pets. However, players do not learn extensive technical skills playing in such commercial game environments. Criteria for truly equitable software must go beyond representation and game scenarios and allow models which empower students to be software designers and have technical and creative control over their own environments.

If girls can access programming processes and compelling environments early, they will be able to make their own creations, construct significant objects and environments which reflect the way girls think, represent their values and opinions, and lead to the formation of new knowledge. In addition, such experience will enable them to participate in the currently male-dominated software industry as a generation of girls becomes technological creators [3]. If we can broaden the participation of all middle school girls with techniques and possibilities that they can relate to their actual, everyday lives and can expand upon in a virtual environment, this will rearrange the dynamics in the creative terrain of computational media. Indeed, the eventual aim of gender equity research should emphasize the diverse range of interests and preferences for all students, so that through good design, creative environments, and customization, we are able to create experiences which will ultimately support a broad spectrum of learners.

The purpose of our project is to build a successful software environment to address this imbalance. The goal is to develop an engaging way to teach computer programming to middle school girls in a scalable, approachable manner that appeals to girls' sense of communication, curiosity, and play. Specifically, this project aims to study how the social design of a programming software environment affects girls' motivations to program, their ability to learn programming concepts, and the extent to which girls' overall confidence and self-efficacy levels are affected by the RAPUNSEL environment.

Specifically, the RAPUNSEL project aims to encourage girls to become familiar with the sort of programming language (in particular Java) that they will encounter several years later, in high school. This is done in the context of a dance game: Players teach animated characters to dance. A large part of the "game mechanic" is the gathering and modification of code that adds to the dance repertoire of the characters. In the course of this, the players not only learn, but also have fun and form a social community of fellow programmers. Ultimately, it is our hope that such a 'self-teaching' environment, in which girls are motivated to learn Java programming incrementally through a pro-active game experience that allows them to master fundamental programming concepts, will lead them to transfer this knowledge to situations outside of our project.

We note that categorizing 'girls' together in a group is a potentially problematic task for which there must be a clear reason, such as working overall for gender equity issues science and technology. Attempting to create something for 'girls' as a category obviously navigates a dangerous border zone between personal, specific, lived experience, and over-generalization. 'Girls' are as diverse in their interests, abilities, and tastes as any other category of people (e.g., 'students' or 'the French'). In gender research in the games industry, designers must be able to work towards gender equity without falling into stereotyping traps, realizing the inherent breadth and contradictions of categorization. The goal of this approach, then, is to design for a multiplicity of experiences, parts of which could be co-opted or rebuilt entirely by the users. Focusing on a few of the niche interests expressed by girls involved in this work may help diversify all kinds of game goals and address numerous types of play styles. Therefore, one way to address designing for girls is to design for a multiplicity of play styles while providing diverse thematic content.

2 Approach

The approach to creating the game has been to posit a virtual world that contains engaging non-player characters. We have been working with 12-13 year old girls in the New York City area, who have been operating as design consultants. As we develop various aspects of the game world and the user interface, we continually go to these design consultants for informal feedback, in a continual cycle of testing and refinement. In particular, we have found it extremely useful to work with girls in computer clubs in Brooklyn, Harlem, and the Bronx.

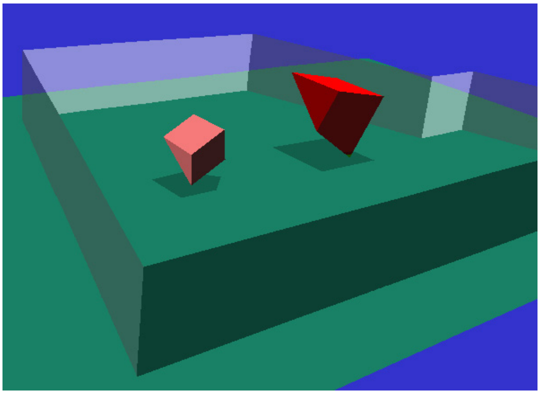
3 Tasks

The major tasks of the RAPUNSEL project have fallen into three broad areas: *(i)* animation design, *(ii)* game-play / interface / feedback, and *(ii)* assessment and measurement. Roughly speaking, these tasks have been divided among the three authors (Perlin/Animation, Flanagan/Game-design, Hollingshead/Assessment), with much of the work implemented by graduate students at New York University and Hunter College. Each of these areas needs to be described in somewhat more detail.

3.1 Animation Design

A key aspect of the game design is that there is an "always running" simulation world. Characters express their personality through movement and of course through choice of apparel. The game design assumes a slow, high-latency synchronization between players. For this reason, the actual animation is done locally, with a periodically updated server maintaining world consistency, using a blackboard model of shared data representation.

In the course of the project, various approaches to appearance and movement were tried for the animated characters and their environment. Our first, relatively abstract, characters were well received by adults and small children, but were not considered "cool" enough by the target group:



We subsequently did systematic testing with our design consultants to find out what sorts of characters they prefer. We presented them with a wide variety of character model ideas, drawn from our own designs (one set of examples is shown above) as well as from the popular culture. These ideas were variably abstract, cartoonish, "cute", realistic, and so forth. Unfortunately, we discovered a near-universal preference among 12-13 year old girls for the sorts of over-sexualized role models currently represented in popular culture by Britney Spears and Christine Aguilera, and we were reluctant to feed into such stereotypes. Eventually we came up with a working compromise - characters that were attractive and sassy, and that the target audience found to be cool, but who were not overly sexualized. One such character is visible in the screen-shot of the next section.

3.2 Game-Play/Interface

Game play in RAPUNSEL consists of several essential components. We want the core of the experience to be the creation, exchange and modification of code, but we also want to establish early within the players a sense of connection and identification with the characters. For this reason, several components were added to game play, including and a separate interaction space for players to customize the appearance of their character, and a private area for players to practice dance moves with their character.

One thing we found decisively in our work with the design consultants is that movement wins out over conversation. Players want to see that characters move, and like to have control over those movements. This turned out to be a significant motivator.

Interface design was a challenge, largely in the sense that it lay at the core of several research questions. Specifically, what is the proper way to introduce code, and how "smart" should a smart text editor be? At one extreme, players can be given a plain text window for editing Java code. But this extreme approach defeats the purpose of a graceful and fun introduction to coding - one impasse caused by a misplaced semicolon and the potential young programmer is turned off to the whole experience.

At the other extreme, a code window which hides all type declarations and semicolons and many other semantic and syntactic elements from the user's view does not serve the purpose of introducing the Java language. Our approach has been to treat this dialectic as a subject for formal testing. We have been building a smart text editor with switches in it, so that various levels of simple to sophisticated presentation are possible.

The interface consists of a set of code and chat windows transparently overlaid onto a screen-filling game play area. After much experimentation we settled on the Torque game engine for the implementation platform. The final decision was based on performance, and the feature set of the game simulator itself, on ease and flexibility of implementing customized text overlays, and on the ease with which we could port our character animation component. Some views of the game with text and graphics overlays are shown below:

3.3 Code Exchange

In order to build a community of game players, the game design encourages - and rewards - players for sharing code with each other in order to collaboratively build character movement and abilities. This promotes a sense of familiarity with code, and encourages players to think of the code itself as something to discuss and to understand together. Below is a set of snapshots from a player/player interaction in which one player is sharing code with another:

1. Player 1 sees that player 2 is doing a cool nove called "doRoll" (in the thought bubble).

2. Player 2 asks, "hey, can I get your doRoll code?"

3. Player 1 Agrees

4. Player 1 highlights move in library and clicks send

5. List of players pops up for player to choose from; choose by clicking on player name.

6. Code appears in Player 2's library

3.4 Assessment

Assessment consists of two phases: a pre-test and a post-test. The pre-test is designed to establish a base level; it is important o measure the level of knowledge, conceptual sophistication, and attitude toward computers and programming that players of the RAPUNSEL game had before encountering the project, in order to properly assess the impact of playing the game upon those factors.

Assessment focuses on the following general questions: *(i)* what game play and goals do girls care about? *(ii)* how much do social status / personality matter? *(iii)* is learning being properly scaffolded?

The term *scaffolding* is used in education to denote the appropriate exposure to new material at each step in the learning of new content or concepts. Proper scaffolding is key to the success of a game-based learning project. If subjects are introduced two quickly or in a confusing order, then players will become frustrated and lose interest. Conversely, if the same concept is dwelt upon for two long, then players will become bored, and again they will lose interest. we use formal assessment to measure the effect of order and rate of advancement for essential programming concepts, including variables, conditional expressions, loops and procedures.

For example, is it better to introduce procedures without arguments early on, and then introduce procedures with arguments only after players have become familiar with variables? Similarly, is it better to introduce variable assignment together with the use of variables, or should this more advanced concept be introduced much later?

These are not questions that can be answered *a priori*, but only through carefully designed assessment protocols with an actual population of kids.

4 Conclusions

RAPUNSEL represents an incremental approach to wide-scale cultural change in gender equity issues in math and science. We are conducting this research to afford cultural diversity, individual learner diversity, and cultural change through a relevant and useful programming environment. The program's goals and open-ended architecture will ultimately encourage others to build communities and artifacts that matter to all individuals and groups.

Not only does RAPUNSEL address issues relating to gender equity and the digital divide, but we also believe that collaborative learning software such as RAPUNSEL will help to facilitate a more fundamental change in society: increasing computer literacy. Computer literacy for all is imperative for any modern society to maintain a diverse, globally engaged workforce of scientists, engineers and well-prepared citizens. This literacy must include computer programming and computer science fundamentals: and involve both reading (using existing computer applications) and writing (making one's own applications). Literacy, however, does not simply involve technical expertise. Literacy is a widespread and socially engaged system of skills, capabilities, and creativity formed in the context of a material support [11]. Therefore we must not only focus on the material support (software environments) but also on the social: the values, interests, motivations, and community practices of learners. Our project takes small steps towards big changes by addressing the unmet needs of middle school-age girls, but we envision a much larger and more fundamental application for the knowledge that will be gained from this project.

References

1. AAUW. (1998). *Gender Gaps: Where Schools Still Fail Our Children*. American Association of University Women Education Foundation, New York.
2. AAUW. (2002). “Education and Software Games: Rethinking the “Girls' Game.” *The Jossey-Bass Reader on Gender in Education*. San Francisco: Jossey-Bass, 345 – 360.
3. Bernstein, D. (1991). “Comfort and Experience with Computing: Are They the Same for Women and Men?” *SIGCSE Bulletin* 23, 1991, 57 -60.
4. Beyer, S. & Bowden, E. (1990). “Gender Differences in the Accuracy of Self-Evaluations of Performance.” *Journal of Personality and Social Psychology*, 59, 960-970.
5. Brandon, D. P., & Hollingshead, A. B. (1999). “Collaborative Learning and Computer-Supported Groups.” *Communication Education,* 48, 109-126.
6. Bruckman, A. & Resnick, M. (1995). “The MediaMOO project: Constructionism and Professional Community.” *Convergence*, 1(1), Spring.
7. Bruckman, A., Jensen, C. & DeBonte, A. (Jan 2002) “Gender and Programming Achievement in a CSCL Environment.” Long talk. *Proceedings of CSCL 2002*. Boulder, CO.
8. Carr, H. H. (1992). “Factors That Affect User-Friendliness in Interactive Computer Programs.” *Information and Management*, 22, 137-149.

9. Carter, J. & Jenkins. T. (June 1999). “Gender and Programming: What's Going On?” Bill Manaris, ed. *Proceedings of the 4th Annual SIGCSE/SIGCUE Conference on Innovation and Technology in Computer Science Education (ITiCSE'99)*, 1-4, June 1999. ACM SIGCSE/SIGCUE.
10. Commission on the Advancement of Women and Minorities in Science, Engineering, and Technology Development (CAWMSET). (Sept 2000) *Land of Plenty: Diversity as America’s Competitive Edge in Science, Engineering and Technology.*. Available online at http://www.nsf.gov/od/cawmset/report.htm
11. diSessa, A. (2000). *Changing Minds: Computers, Learning, and Literacy.* Cambridge: MIT Press.
12. Elliot, C. Lester, J., & Rickel, J. (1997). “Integrating Affective Computing into Animated Tutoring Agents.” In *Proceedings of the IJCAI Workshop on Animated Interface Agents: Making Them Intelligent*, Nagoya, Japan, 113-121.
13. Furger, R. (1998). *Does Jane Compute? Preserving Our Daughters’ Place in the Cyber Revolution*. New York: Time Warner.
14. Haller, S. & Fossum, T. (1998). “Retaining Women in CS with Accessible Role Models.” *Proceedings of SIGCSE’98 Conference,* Atlanta.
15. Honey, M., Moeller, B., Brunner, C., Bennett, D., Clements, P., & Hawkins, J. (2002). “Girls And Design: Exploring the Question of Technological Imagination.” *The Jossey-Bass Reader on Gender in Education*. San Francisco: Jossey-Bass, 329-34.
16. Lynch, D. (2000). “Purple Moon Setting.” *ABC News Internet Edition.* ttp://abcnews.go.com/sections/ tech/WiredWomen/wiredwomen990309.html
17. McLester, S. (Oct 1998). “Girls and Technology: What's the Story?” *Technology & Learning*, 19:3, 18.
18. Moller, L. C., Hymel, S., & Rubin, K. (April 1992). “Sex Typing in Play and Popularity in Middle Childhood.” *Sex Roles: A Journal of Research*. 26:7-8, 331- 354.
19. Orenstein, P. (1995). *Schoolgirls: Young Women, Self-Esteem, and the Confidence Gap.* New York: Doubleday & Company.
20. Pearl, A., Pollock, M., Riskin, E., Thomas, B., Wolf, E., & Wu, A. (1990). “Becoming a Computer Scientist.” *Communications of the ACM*, 33:11, 47-57.
21. Smith, T. E. (1992).”Gender Differences in the Scientific Achievement of Adolescents: Effects of Age and Parental Separation.” *Social Forces,* 71, 469-484.
22. Taylor Anderson, B. (1993). “How Can Middle Schools Get Minority Females in the Math/Science Pipeline?” *Education Digest.* 59:2, 39.
23. U.S. Bureau of Labor Statistics. (2002). *Report 3915*. http://www.bls.gov/rofod/3915.pdf; Accessed 2002. http://www.bls.gov/
24. Wigfield, A. & Eccles, J. S. (1994). “Children's Competence Beliefs, Achievement Values, and General Self-Esteem Change Across Elementary and Middle School.” *Journal of Early Adolescence 14*, 107-138.
25. Wilson, B. C. & Shrock, S. (2001). “Contributing to Success in an Introductory Computer Science Course: A Study of Twelve Factors.” *ACM SIGCSE Bulletin, Proceedings of the thirty second SIGCSE technical symposium on Computer Science Education, 33*(1), 184-188.

Automatic Conversion from E-Content into Virtual Storytelling

Kaoru Sumi[1] and Katsumi Tanaka[2]

[1] National Institute of Information and Communications Technology,
Interactive Communication Media and Contents Group,
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0289, Japan
Kaoru@nict.go.jp
http://www2.nict.go.jp/jt/a133/indexe.html
[2] Kyoto University, Graduate School of Informatics,
Yoshida Honmachi, Sakyo, Kyoto 606-8501, Japan
ktanaka@i.kyoto-u.ac.jp

Abstract. This paper describes a virtual storytelling medium, called *Interactive e-Hon*, for helping children understand content from the Web or other electronic devices. It works by transforming electronic content into an easily understandable "storybook world." In this world, easy-to-understand content is generated by creating 3D animations that include content and metaphors, and by using a child-parent model with dialogue expression and a question-answering style comprehensible to children.

1 Introduction

We are awash in information flowing from the World Wide Web, newspapers, and other types of documents, yet the information is often hard to understand; laypeople, the elderly, and children find much of what is available incomprehensible. Thus far, most children have missed opportunities to use such information, because it has been prepared by adults, for adults. The volume of information specifically intended for children is extremely limited, and it is still primarily adults who experience the globalizing effects of the Web and other global networks. The barriers for children include difficult expressions, prerequisite background knowledge, and so on. Our goal is to remove these barriers and build bridges to facilitate children's understanding and curiosity. In this research, we are presently considering the applicability of systems designed to help children's understanding.

This paper describes a virtual storytelling medium, called Interactive e-Hon, for helping children understand difficult topics. The Japanese word *hon* means "book", while *ehon* means "picture book". Interactive e-hon works by transforming electronic content into an easily understandable "storybook world." It uses animation to help children understand the content. Visual data attract a child's interest, and the use of concrete examples like metaphors facilitates understanding, because each person learns according to his or her own unique mental model [1][2], formed according to the person's background. For example, if a user poses a question about something, a system that answers with a concrete example in accordance with the user's specializa-

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 260 – 269, 2005.

tion would be very helpful. For users who are children, an appropriate domain might be a storybook world. With this approach in mind, our long-term goal is to help broaden children's intellectual curiosity [3] by broadening their world.

Attempts to transform natural language (NL) into animation began in the 1970s with SHRDLU [4], which represents a building-block world and shows animations of adding or removing blocks. In the 1980s and 1990s, more applications [5][6][7] appeared, in which users operate human agents or other animated entities derived from NL understanding. Recently, there has been research on the natural behaviour of life-like agents in interactions with users [8][9][10]. The main theme in this line of inquiry is the question of how to make these agents as human-like as possible in terms of dialogicality, believability, and reliability. On the other hand, our research aims to make content easier for users to understand, as opposed to improving agent humanity. WordsEye [11] is a text-to-scene generation system that includes special data. In contrast, our system generates animations but not scenes. Thus, we have pursued this line of inquiry because little or no attention has been paid to media translation from content with the goal of improving users' understanding.

2 Interactive e-Hon

Interactive e-Hon is a kind of word translation medium that provides expression through the use of 3D animation and dialogue explanation in order to help children understand Web content or other electronic resources, e.g., news, novels, and essays. For a given content, an animation plays in synchronization with a dialogue explanation, which is spoken by a voice synthesizer.

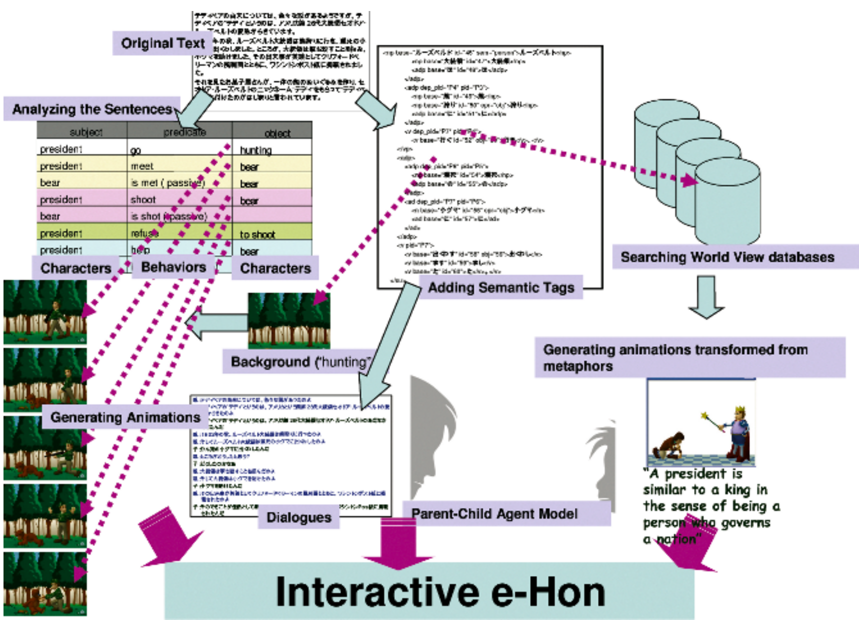


Fig. 1. Overview of Process Flow for Interactive e-Hon

The main points of our approach are to show (1) animations of behavioral semantics, i.e., who is doing what, etc.; and (2) dialogues between the shadows of a parent and a child, represented as cartoon characters through a simple visual interface and intended for situations in which a parent and child use the system together. The main functions of Interactive e-Hon are (1) transforming NL into animation word by word, (2) generating dialogue from the NL of a content, (3) presenting the content by using parent and child agents, and (4) explaining concepts through the use of animations transformed from metaphors.

Figure 1 shows the process flow for Interactive e-Hon. This processing is based on text information containing semantic tags that follow the Global Document Annotation (GDA)[1] tagging standard, along with other, additional semantic tags. Tags with several semantic meanings for every morpheme, such as "length," "weight," "organization," and so forth, are used. To provide normal answers, the system searches for tags according to the meaning of a question. To provide both generalized and concretized answers, the system, after searching the tags and obtaining one normal answer, generalizes or concretizes the answer by using ontologies. Recently, the Semantic Web [12] and its associated activities have adopted tagged documentation. Tagging is also expected to be applied in the next generation of Web documentation. Note that our system uses the Japanese language for text.

In the following subsections, we describe the key aspects of Interactive e-Hon: the information presentation model using parent and child agents, the transformation of electronic content into dialogue expressions, the transformation of electronic content into animation, and the expression of conceptual metaphors by animation.

2.1 Parent-Child Agent Model

We believe that (1) an easy, natural, unforced viewing style and (2) a feeling of assimilation into content are important elements in encouraging users to accept and become interested in the content. For example, with a guided agent, however eagerly it talks to a user, the user may not be interested and may no longer want to watch. We think that this is a kind of forced style, similar to how the more parents tell children to study, the less they feel like studying. Furthermore, for users to be interested in content, it must provide some form of assimilation for users. If users feel that a virtual world is completely separate from themselves, they will lose interest in it.

Our system has agents that mediate a user's understanding through intelligent information presentation. In the proposed model, a parent agent (mother or father) and a child agent have a conversation while watching a "movie" about content, and the user (or users, in the case of a child and parent together) watches the agents. The agents take the form of moving shadows of the parent and child. The shadow of the child agent represents the shadow of the real child user, and the shadow of the parent agent represents the shadow of the child's real parent. These shadows appear on the system's screen and move freely without the user's intervention, similarly to the way Peter Pan's shadow moves independently. The shadows enhance the feeling of as-

[1] http://i-content.org/GDA
Internet authors can annotate electronic documents with a common, standard tag set, allowing machines to automatically recognize the semantic and pragmatic structures of the documents.

similation between the real world (users) and the virtual world (content). The user's assimilation with the content should lead to a feeling of affinity.

There are two kinds of agents: avatars, which represent the self, and guides or actors, which represent others. Avatars are more agentive, dialogical, and familiar than guides or actors [13]. Thus, we designed the avatars in Interactive e-hon so that users would feel familiarity and affinity with the agents, helping them gain a better understanding of content.

2.2 Transformation from Content into Dialogue Expressions

To transform content into dialogue and animation, the system first generates a list of subjects, objects, predicates, and modifiers from the text information of the content. It also attempts to shorten and divide long, complicated sentences.

Then, by collecting these words and connecting them in a friendly, colloquial style, the system generates conversational sentences. In addition, it reduces the level of repetition for the conversational partner by changing phrases according to a thesaurus. It prepares explanations through abstraction and concretization based on ontologies, meaning that it adds explanations of background knowledge. For example, in the case of abstraction, "Antananarivo in Madagascar" can be changed into "the city of Antananarivo in the nation of Madagascar," which uses the ontologies, "Antananarivo is a city" and "Madagascar is a nation." Similarly, in the case of concretization, "woodwind" can be changed into "woodwind; for example, a clarinet, saxophone, or flute." These transformations make it easier for children to understand concepts.

In the case of abstraction, the semantic tag "person" adds the expression, "person whose name is"; "location" adds "the area of" or "the nation of"; and "organization" adds "the organization of." In the case of concretization, if a target concept includes lower-level concepts, the system employs explanations of these concepts.

2.3 Transformation from Content into Animation

Interactive e-Hon transforms content into animation by using the word list described in the previous subsection. In an animation, a subject is treated as a character, and a predicate is treated as an action. An object is also treated as a character, and an associated predicate is treated as a passive action. One animation and one dialogue are generated for each list, and these are then played at the same time.

Many characters and actions have been recorded in our database. At present, the system has 18 characters, 67 behaviors, and 31 backgrounds, but the database is still being expanded. A character or action involves a one-to-many relationship. Various character names are linked to each character. Various action names are linked to each action, because often several different names indicate the same action. Actions can be shared among characters in order to prepare a commoditized framework of characters. The representation in an animation comes from real motions, gestures, emphasized behaviors, and cartoon-like deformations.

If there is a word correspondence between the name of a character and a subject or object in the word list, the character is selected. If the subject or object is a company, a government, or a public agency, a man in a suit is selected. If there is no word correspondence, as in the case of the semantic tag "person," the system selects a general person character according to an ontology of characters. As for the background file, it is selected in the priority order of the semantic tags "location," "pur" (purpose), and "sbj" (subject).

2.4 Searching and Transformation of Metaphors into Animation

If a user does not know the meaning of a term like "president," it would be helpful to present a dialogue explaining that "a president is similar to a king in the sense of being the person who governs a nation," together with an animation of a king in a small window, as illustrated in Figure 2. People achieve understanding of unfamiliar concepts by transforming the concepts according to their own mental models [1][2]. The above example follows this process.

The dialogue explanation depends on the results of searching world-view databases. These databases describe the real world, storybooks (with which children are readily familiar), insects, flowers, stars, and so on. The world used depends on a user's curiosity, as determined from the user's input in the main menu. For example, "a company president controls a company" appears in the common world-view database, while "a king reigns over a country" appears in the world-view database for storybooks, which is the target database for the present research. The explanation of "a company president" is searched for in the storybook world-view database by utilizing synonyms from a thesaurus. Then, the system searches for "king" and obtains the explanation: "A company president, who controls a company, is similar to a king, who governs a nation." If the user asks the meaning of "company president," the system shows an animation of a king in a small window while a parent agent, voiced by the voice synthesizer, explains the meaning by expressing the results of the search process.

In terms of search priorities, the system uses the following order: (1) complete correspondence of an object and a predicate; (2) correspondence of an object and a predicate, including synonyms; (3) correspondence of a predicate; and (4) correspondence of a predicate, including synonyms. Commonsense computing [14] is an area of related research on describing world-views by using NL processing. In that research, world-views are transformed into networks with well-defined data, like semantic networks. A special feature of our research is that we directly apply NL with semantic tags by using ontologies and a thesaurus.

3 Application

We tried converting four actual examples of content from the Web, newspapers, and novels. For example, we transformed the content, "the origin of the *teddy bear*'s name," from a Web source[2] into an animation and a dialogue (Figure 2).

3.1 Transformation of Web Content into Dialogue

As described above, the system first generates a list of subjects, objects, predicates, and modifiers from the text information of content; it then divides the sentences in the text. For example, it might generate the following lists from the long sentences shown below, using the following abbreviations: MS: modifier of subject; S: subject; MO: modifier of object; O: object; MP: modifier of predicate; P: predicate.

[2] http://www.jteddy.net/topics/bear.html

In this case, Interactive e-Hon is explaining the concept of a “president” by showing an animation of a king. The mother and child agents talk about the content. The original text information can be seen in the text box above the animation. The user can ask questions directly in the text box.

The original text:

“President Roosevelt went bear hunting and met a dying bear in autumn of 1902. However, the President refused to shoot and kill the bear, and he helped it instead. Along with a caricature by Clifford Berryman, the occurrence was carried by the *Washington Post* as a heartwarming story.”

The following is a dialogue explanation for this example:

Parent Agent: President Roosevelt went bear hunting. Then, he met a small, dying bear.

Child Agent: The President met a small bear who was likely to die.

A real child: *What is a president, mummy? (Then, his mother operates the e-Hon system by clicking the word. Here, an animation using the retrieved metaphor is played.)*

Parent agent: A president is similar to a king as a person who governs a country (with the king animation in a small window).

A real parent: *A president is a great man, you know?*

Parent Agent: But, what do you think happens after that?

Child Agent: I can’t guess. Tell me the story.

Parent Agent: The President refused to shoot and kill the bear. And, he helped it instead.

Child Agent: The President assisted the small bear.

Parent Agent: The occurrence was carried by the Washington Post as a heartwarming story,

with a caricature by Clifford Berryman.

Child Agent: The episode was carried by the newspaper as a good story.

Fig. 2. A sample view from Interactive e-Hon

(Original Sentence 1)

“It is said that a confectioner, who read the newspaper, made a stuffed bear, found the nickname “Teddy,” and named it a “teddy bear.”

(List 1)

- S: confectioner; MS: who read the newspaper; P: make; O: stuffed bear.
- S: confectioner; P: find; O: nickname “Teddy”; MO: his.
- S: confectioner; P: name; MP: “Teddy bear”.
- S: it; P: say.

(Original Sentence 2)
"But, the president refused to shoot the little bear, and he helped it instead."
(List 2)
- S: president; P: shoot; O: little bear.
- S: president; P: refuse; O: to shoot the little bear.
- S: president; P: help; O: little bear.

The system then generates dialogue lines one by one, putting them in the order (in Japanese) of the subject's modifier, the subject, the object's modifier, the object, the predicate's modifier, and the predicate, according to the line units in the list. To provide the characteristics of real storytelling, the system uses past tense and speaks differently depending on whether the parent agent is a mother or father.

Sometimes the original content uses reverse conjunction, as with "but" or "however," as in the following example: "But, what do you think happens after that?"; "I can't guess. Tell me the story." In such cases, the parent and child agents speak by using questions and answers to spice up the dialogue. Also, at the end of every scene, the system repeats the same meaning with different words by using synonyms.

3.2 Transformation of Web Content into Animation

In generating an animation, the system combines separate animations of the subject as a character, the object as a passive character, and the predicate as an action, according to the line units in the list.

For example, in the case of Original Sentence 2 above, first

- Pr esident (character) shoot (action), and
- little bear (character; passive) is shot (action; passive)
  are selected. After that,
- president (character) refuse (action)
  is selected. Finally,
- president (character) help (action), and
- little bear (character; passive) is helped (action; passive)
  are selected.

This articulation of animation is used only for verbs with clear actions. For example, the be-verb and certain common expressions, such as "come from" and "said to be" in English, cannot be expressed. Because there are so many expressions like these, the system does not register verbs for such expressions as potential candidates for animation.

4 Experiment

We conducted experiments using real subjects to examine whether Interactive e-Hon's expression of dialogue and animation was helpful for users.

In the first experiment, we had 18 children, ranging from 7 to 15 years old, compare an original text with its expression via Interactive e-Hon. We showed four different examples of content from the Web, newspapers, and novels. This experiment was conducted through in-person interviews. All of the children participated in this experiment with their parents. The subjects were shown both the original text on paper and its expression through Interactive e-Hon. All of the children responded that Interactive e-Hon was more interesting and easier to learn from as compared to the original text.

In the next experiment, we used the example of "the origin of the teddy bear's name." Three types of content were presented to the users and evaluated by them: the original content read by a voice synthesizer (content 1), a dialogue generated by Interactive e-Hon and read by a voice synthesizer (content 2), and a dialogue with animation generated by Interactive e-Hon and read by a voice synthesizer (content 3). The subjects were Miss T and Miss S, both in their twenties; child K, five years old; and child Y, three years old. Both women understood content 2 as a dialogue but found content 1 easier to understand because of its compaction. They also thought content 3 was easier to understand than content 2 because of its animation. T, however, liked content 1 the best, while S favored content 3. As T commented, "Content 1 is the easiest to understand, though content 3 is the most impressive." In contrast, S commented, "Content 3 is impressive even if I don't hear it in earnest. Content 1 is familiar to me like TV or radio." She also noted, "The animation is impressive. I think the dialogues are user-friendly and may be easy for children to understand."

K, who was five years old, said that he did not understand content 1. He initially felt that he understood content 2 a little bit, but he could not express his own words about it. He found content 3, however, entirely different from the others, because he felt that he understood it, including the difficult Japanese word *kobamu*, meaning "refuse." Child Y, who was three years old, showed no recognition of contents 1 and 2, but he seemed to understand content 3 very well, as he was able to give his thoughts on the content by asking (about President Roosevelt), "Is he kind?".

In this experiment, we observed a clear difference in the results for the adults and the children, despite the limited number and age range of the subjects. The two adults felt the content presentation style with dialogue and animation was easier than that using dialogue only. One of the adults, however, felt that hearing the original content read was the best approach for understanding because of its compaction, while the other felt that the content presentation style with dialogue and animation was the best approach. In contrast, the children did not understand when they heard either the original content or the dialogue-based content read. They only understood the content presentation style with dialogue and animation.

5 Discussion

Interactive e-Hon's method of expression through dialogue and animation is based on NL processing of contents. For dialogue expression, the system generates a plausible, colloquial style that is easy to understand, by shortening long sentences and extracting the subject, objects, predicate, and modifiers from it. For animation expression, the system generates a helpful animation by connecting individual animations selected for the subject, objects, and predicate. The result is expression through dialogue with animation that can support a child user's understanding, as demonstrated by our experiment using real subjects.

If the natural language processing (NLP) is successful and there are animations appropriate for corresponding words, the system can exchange original text into dialogue and animation. Reference terms (e.g., "it," "that," "this," etc.) and verbal omission of a subject are open problems in NLP and remain as issues in our system. As a tentative solution, we have manually embedded word references in GDA tags. A fully automatic process knowing which words to reference will depend upon further progress in NLP.

When translating a word into an animation, ambiguity may be a problem. This can be solved by using a dictionary that can automatically detect the meaning of a word enumeration pattern. There is also the problem of ambiguity in the animations. We are now verifying whether the behaviors that we have prepared are consistently recognized as corresponding to the intended words. We will therefore have to regulate the word enumeration patterns.

In the process of registering character data and the corresponding words, or an action and its corresponding words, which are one-to-many correspondences, certain groups of words that are like new synonyms are generated via the 3D content. These groups of synonyms are different from NL synonyms, and new relationships between words can be observed. This feature can be considered for potential application as a more practical thesaurus based on 3D content, as opposed to an NL thesaurus.

In the experiment, we observed effective support for understanding with the system and the possibility of appreciation for learning new words. We have concluded that content that is within a user's background knowledge is easier to understand by regular reading, as in the case of the adults in this experiment. In contrast, for content outside a user's background knowledge, animation is expected to be very helpful for understanding, as in the case of the children. Further experiments may show that for a given user, difficult content outside the user's background knowledge can be understood through animation, regardless of the user's age.

6 Conclusion

We have introduced Interactive e-Hon, a system for facilitating children's understanding of electronic content by transforming it into a "storybook world." We have conducted media transformation of actual Web content and demonstrated the effectiveness of this approach via an experiment using real subjects. We have thus shown that Interactive e-Hon can generate satisfactory explanations of concepts through both animation and dialogue that can be readily understood by children.

Interactive e-Hon could be widely applied as an assistant to support the understanding of difficult contents or concepts by various kinds of people with different types of background knowledge, such as the elderly, people from different regions or cultures, or laypeople learning about a difficult field.

In our future work, we will further evaluate the effectiveness of the system in relation to human understanding. We also plan to expand the animation and word databases and to apply Interactive e-Hon to several other kinds of content.

References

1. Philip N. Johnson-Laird: Mental Models, Cambridge: Cambridge University Press. Cambridge, Mass.: Harvard University Press (1983).
2. D. A. Norman, *The Psychology of Everyday Things*, Basic Books (1988).
3. Hatano and Inagaki: Intellectual Curiosity, Cyuko Shinsho (in Japanese) (1973).
4. Terry Winograd, Understanding Natural Language, Academic Press (1972).
5. Vere, S. and Bickmore, T.: A basic agent. Computational Intelligence, 6:41-60 (1990).
6. Richard A. Bolt: “Put-that-there”: Voice and gesture at the graphics interface, International Conference on Computer Graphics and Interactive Techniques archive, Proceedings of the 7th annual conference on Computer graphics and interactive techniques, ACM Press (1980).
7. N. Badler, C. Phillips, and B. Webber, Simulating Humans: Computer Graphics, Animation and Control. Oxford University Press (1993).
8. Justine Cassel, Hannes Hogni Vilhjalmsson and Timothy Bickmore: BEAT: the Behavior Expression Animation Toolkit, Life-Like Characters, Helmet Prendinger and Mitsuru Ishizuka Eds., pp. 163-187, Springer (2004).
9. Hozumi Tanaka et al: Animated Agents Capable of Understanding Natural Language and Performing Actions, Life-Like Characters, Helmet Prendinger and Mitsuru Ishizuka Eds., pp. 163-187, Springer, (2004).
10. Stacy Marsella, Jonathan Gratch and Jeff Rickel: Expressive Behaviors for Virtual World, Life-Like Characters, Helmet Prendinger and Mitsuru Ishizuka Eds., pp. 163-187, Springer (2004).
11. Bob Coyne and Richard Sproat: WordsEye: An Automatic Text-to-Scene Conversion System, SIGGRAPH 2001, Proceedings of the 28th Annual Conference on Computer Graphics, Los Angeles, California, USA. ACM (2001).
12. D. Fensel, J. Hendler, H. Liebermann, and W. Wahlster (Eds.) *Spinning the Semantic Web*, MIT Press (2002).
13. Toyoaki Nishida, Tetsuo Kinoshita, Yasuhiko Kitamura and Kenji Mase: *Agent Technology,* Omu Sya (in Japanese) (2002).
14. Hugo Liu and Push Singh: Commonsense reasoning in and over natural language. Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES-2004) (2004).

The Lost Cosmonaut: An Interactive Narrative Environment on the Basis of Digitally Enhanced Paper

Axel Vogelsang[1] and Beat Signer[2]

[1] University of the Arts,
Central Saint Martins College,
Southampton Row, London
a.vogelsang1@csm.arts.ac.uk
http://www.incompatibility.net
[2] Institute for Information Systems,
ETH Zurich, CH-8092 Zurich, Switzerland
signer@inf.ethz.ch

Abstract. The Lost Cosmonaut is an interactive narrative based on digitally enhanced paper. This technology uses an electronic pen to mediate between paper and computer. Thus any actions of the pen on the paper can be captured and manipulated by a computer as well as we can map digitally controlled events onto paper. The story in this narrative environment reveals itself partially through written text and images on the paper surface just as any other printed story. However, additional information in form of digitally controlled outputs such as sound, light and projections can be accessed through interaction with pen and paper. Furthermore the audience is not only supposed to read and otherwise perceive information, we also want them to actively produce content for this environment by writing onto the paper. By doing so they also add content to the database containing the digital output at the same time. Hence we produce a complex multimedia environment that works on three levels: On paper, in a digitally controlled visual and acoustic environment and in the combination of both worlds. Last but not least this environment is an open system, which grows as a collaborative effort over time as each user adds his own entries to paper and database. We argue that using paper as an integrated part of a digital environment is a best-of-both-world approach that opens up new possibilities for producing and perceiving narrative.

1 Preliminary About the Nature of This Project

The Lost Cosmonaut is an art-science collaboration as part of the Artists-in-Labs program initiated by the HGKZ (University of Applied Sciences and Arts Zurich). The motivation was to develop a stand-alone artwork that showcases the possibilities of technologies developed by the Global Information Systems research group (GlobIS) in regards to interactive narrative. It was not in the

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 270–279, 2005.

scope of the project to develop a general framework for narrative environments with digitally enhanced paper. The very restricted time frame of six months put some constraints on the functionality supported by a first prototype and did not allow for a proper series of user testing. There are some efforts made at the moment to further develop the project into a robust installation for general exhibition.

2 Introduction

Paper has been one of the most common media for storing and exchanging narratives for hundreds of years if not for two millennia. With the arrival of personal computing and notions of emergence and virtual environments the days of paper seemed to be running out. Experts reckoned that the arrival of the paperless office was just a question of time. But why has paper defied all these swan songs so far? And why should paper play a role in a digitally controlled environment in the first place? Research all across human sciences such as by the biologists Maturana and Varela [4] or the psychologist J. J. Gibson [2] has shown that human thinking and acting has developed through interaction with the physical and social environment in an intense process of structural coupling. We are not independent agents in a closed environment. We have developed in relationship to the environment while we have shaped it at the same time. Sellen and Harper [7] among others have looked at why paper still is a crucial media in working environments. According to them *"engagement with paper for the purpose of marking or reading is direct and local. In other words, the medium is immediately responsive to executed actions, and interaction depends on physical copresence."* These and other properties relevant to humans are an expression of our existence in physical space and they have formed our relationship with information during the centuries. Thus Sellen and Harper conclude that *"paper tends to find its natural place in workplace activities that are point-of-use activities or that are the kinds of activities we normally think of as key to knowledge work. These are the activities that involve making judgments, solving problems, making sense of information, making plans, or forming mental pictures of information. In other words, these are the activities we have come to think of as getting to grips with information."* As a lot of these mentioned activities do apply to story telling and reading, we conclude that paper as an integrated part of a digital environment can open new opportunities for interactive narrative as it combines the best of both worlds.

3 Background

The integration of paper and computers has been researched as early as the beginning of the 1990s when Pierre Wellner developed his DigitalDesk [9]. Here are some of the more recent projects dealing with computer-paper-integration and/or with embodied interactive story telling.

The *Listen Reader* [1] was mainly aimed at children and the interface consists of a book containing a printed story combined with a chair with integrated speakers. The user sits on the chair while reading the book. Electronic field sensors in the binding sense the proximity of the user's hand to the pages. While turning the pages or passing over the text the hand movements trigger the playback of music samples as well as they influence volume and pitch. Even though the user still has to read the story, the interaction with the book provokes a physical engagement while the relating sound environment adds an emotional layer to the narrative.

Tangible Viewpoints [5] is a multiple viewpoint narrative that allows the use of physical pawns as "handle on the character viewpoints" in an interactive narrative. One or more users can experience a story from different viewpoints by moving the pawns on a sensitive field. As narrative outputs for this environment serve multiple media such as audio, image, video and text that react on the movements of the pawns. It is a very media-rich environment which engages with space and offers physical embodiment of individual story characters. Thereby it claims to heighten the engagement with the story and the identification with the story characters.

The *EdFest* system [6] is an ongoing research project by the GlobIS group aimed at developing a knowledge sharing environment for the Edinburgh festival based on paper-computer integration. It is supposed not only to provide access to information about venues and events, but also to allow tourists to enter and share reviews of events and information about restaurants and bars discovered during their visit. Tourists are provided with a specially designed digitally augmented map and brochure, a digital pen, an earpiece that serves as an audio channel and a wearable computer. A user can access information by moving the pen over the brochure as well as generate additional input by writing onto it. Moreover these comments will be shared with other users, which also has a strong element of narrative to it.

The Lost Cosmonaut aims at combining a lot of the above-mentioned features: It wants to draw from the strengths of pen and paper as traditional and simple tools for reading and writing stories. It will use the inherent appeal of tactile sensation and spatial movement. It wants to enrich the user experience by a layer of digitally controlled outputs in form of audio, video, image and text. It will produce additional involvement by engaging the user into a collaborative process of narration.

4 Interactive Paper

While several projects in the past have sought to bridge the paper and digital divide, recent technological developments make this a realistic option and a number of commercial products are now available using Anoto technology. The technology is based on an almost invisible pattern of infrared-absorbing dots printed on paper and special digital pens which have a camera situated alongside the writing stylus as shown in Fig. 1.

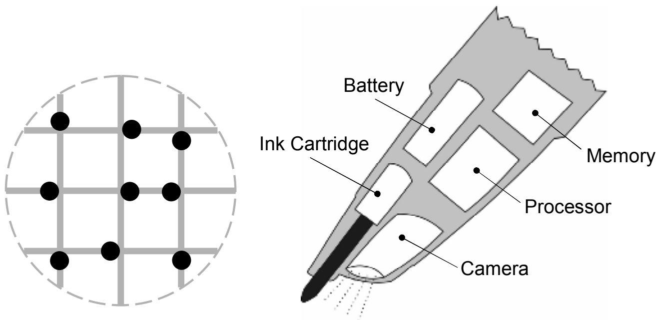


Fig. 1. Anoto technology

The pattern of dots encodes x and y positions in a vast virtual document address space. Camera images are recorded and processed to give up to 100 x and y positions per second and enable a good representation of handwriting to be captured. Several pages of handwriting can be stored in the pen before being transmitted to a PC. Commercial products based on this technology include Nokia's Digital Pen and Logitech's io Personal Digital.

Currently, the focus of commercial products is on information capture for enhanced writing. A number of research projects have also investigated technologies for enhanced reading based on some means of creating links to digital materials within printed documents. Many systems use printed visible marks on paper such as standard barcodes or specially printed patterns. While visible encodings have the advantage of making links obvious to users, they can also be extremely disruptive to reading, especially if many links are present.

The GlobIS research group developed iServer, a cross-media information platform that allows any form of media to be integrated through a resource plug-in mechanism [8]. So far plug-ins for interactive paper (iPaper), HTML documents, movies, sounds and still images have been developed. These allow not only entire resources to be linked, but also elements within resources as specified through appropriate selector mechanisms. In the case of interactive paper, selectors are defined as active areas within physical document pages.

In the Lost Cosmonaut installation, the pen is used as a writing and interaction device, thereby combining the approaches of enhanced writing and enhanced reading. Since iServer is independent of any specific technology for interactive paper, we are able to use Anoto functionality both for the capture of handwriting and the activation of links within a page. However, it is important to stress that in order to do this, we use pens specially modified by Anoto rather than an off-the-shelf product.

5 The Story

We decided against a linear or a forked path that would lead to a closed story. Instead we produced a database of smallest narrative units. The audience is provided with a theme rather than with a coherent story in order to keep the environment open for individual imagination and user contribution.

The theme is based on stories surrounding cosmonauts being either secretly used as guinea-pigs or otherwise lost in space. The emotional and narrative potential was more important to us than the facts behind the stories. There is an element of mystery, which leaves sufficient gaps for the audience to fill in. Space travel has been a generator of fantasy and speculation for a long time. The stars themselves have a long history as an oracle of human future, personal success and love issues. The cosmonaut theme is laden with symbolism and metaphors, waiting to be exploited. It also relates to the time when digital technology slowly started to take over, which together with the nostalgia surrounding the early days of space travel did fit nicely into our idea of combining a very old interface with latest technology.

6 The Setup

The basic setup for The Lost Cosmonaut is a desk on which one user at a given time interacts with objects made of digitally enhanced paper using an electronic pen as shown in Fig. 2. These objects contain pre-produced texts and/or images. Besides reading and looking at those information the user is supposed to point his pen at text and pictures in order to trigger digitally controlled outputs. Those outputs manifest as audio from surrounding speakers as well as video, animation, image, text, projected onto a screen in front of the user.

6.1 Database Content

For the content we gathered film footage and images surrounding space travel. Additionally we used sounds and short speech recordings. These are our smallest narrative units that provide the user with a backdrop for his own personal interpretation of the story while he interacts with our installation.

Another part of content is provided by the audience itself who is invited to write and draw on the interactive objects as we will explain in detail later. These

Fig. 2. The Lost Cosmonaut setup

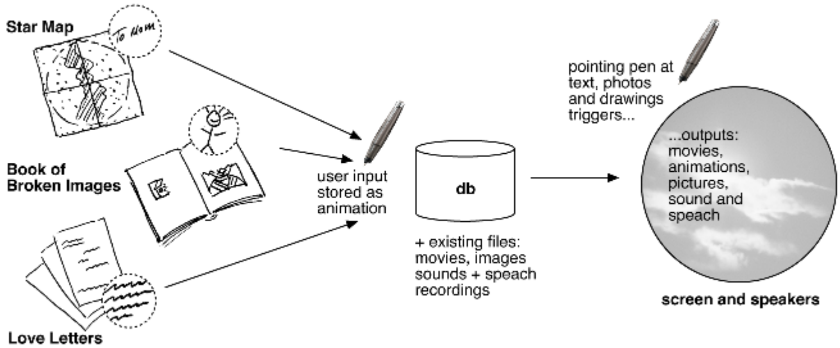


Fig. 3. Content structure

acts of writing and drawing will obviously change the content of the physical object. And in addition this user input will be captured as an animation and will enter the database. Triggered by user interaction these animations will be replayed on the screen at a later time in combination with sounds and speech. So the user's input will reappear as output in a different context.

It is also very important that the whole system, physical objects and database that is, grow over the time of an exhibition. There is no reset of the database neither do we intend to give every user fresh paper objects. Thereby the content builds up over multiple user sessions and the narrative turns into a collaborative effort as outlined in Fig. 3.

6.2 The Paper Objects

The Book of Broken Images: The first object is a photo album presented in Fig. 4. There are photos of people, houses, a party, a starry night, a man in uniform, etc. It could be somebody's family album. The mysterious thing though is that

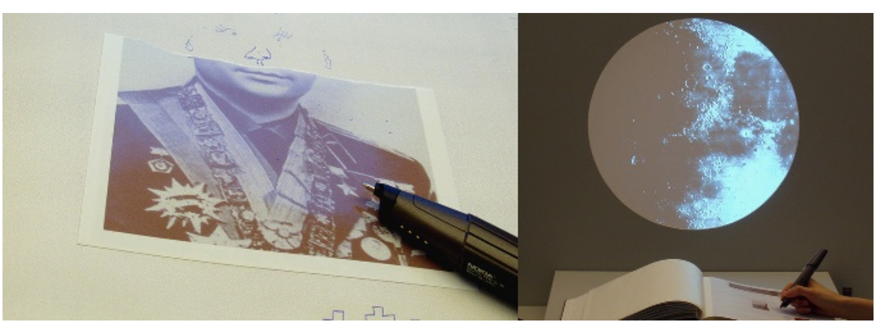


Fig. 4. The Book of Broken Images

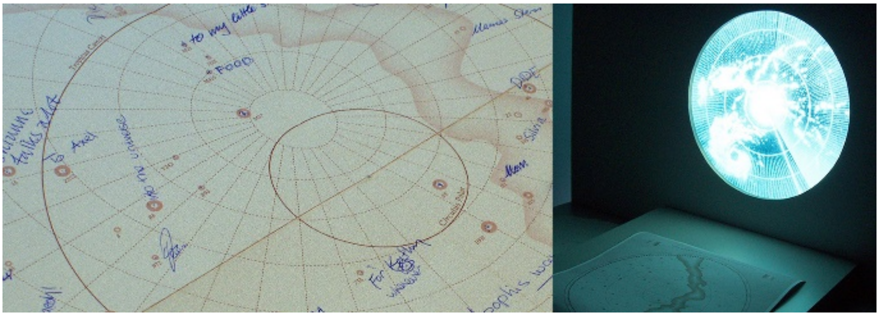


Fig. 5. The Star Map and relating mood

one half of each image is torn off. Those half images are printed on digitally enhanced paper. When the user points with the pen at them, the screen shows different images or films: a camera walking through empty rooms, a desert, the moon surface, all supported by relating sounds.

Each object implies a specific interaction. With the photo album the idea is that the audience should start to imagine the other part of the photo and to fill in the gaps by drawing with the pen. And each drawing itself turns into a sensitive field that is connected with objects in the database. If the user moves the pen onto them he will again be presented with audio or video output. And as we've mentioned before, everything that is drawn is captured and stored in the database.

The Star Map: The second object, shown in Fig. 5, resembles an ancient star map. When the user ticks the pen onto one of the stars, the environment reacts with various outputs coming from the database, which also contains animated versions of the drawings that the users have left in the book of broken images.

Also the user is intended to dedicate a star to somebody. And as with the book of broken images, the dedications added by the users are captured and added to the database while their physical manifestations on the paper turn into active areas, which trigger an output whenever the pen touches.

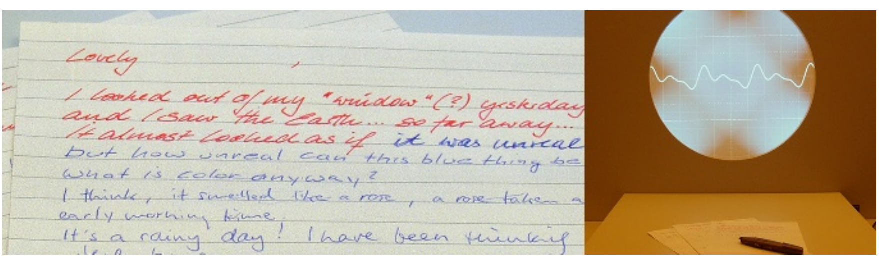


Fig. 6. The Love Letters and relating mood

The Love Letters: They are actually not more than fragments of letters, started but never finished as indicated in Fig. 6. Similar to the other two objects the user interacts: He points the pen at written texts and the system reacts with outputs. He fills in gaps in the letters and thereby produces more content for the database as well as more active areas on the paper which themselves can be used to trigger audio and visuals.

Each of the three objects is additionally supported by a basic theme composed of sound, visual and lighting, which we call mood. This mood is constantly playing/showing as a background when there is no user activity but it is also giving direct feedback on whatever inputs. Part of the mood of the Love Letters is a pulsating heartbeat on an oscilloscope. When the user writes, this heartbeat accelerates. RFID tag stickers on the back of each object in combination with an RFID reader on the middle of the table activate those moods.

7 Technical Infrastructure

The interactive paper platform (iPaper) supports enhanced reading as well as enhanced writing. Based on a set of abstract interfaces, the interactive paper platform is very flexible in supporting various forms of input devices. For the Lost Cosmonaut interactive narrative, a Nokia Digital Pen was the primary input device next to some other sensors, such as Radio Frequency Identifier (RFID) tags, which were used for document tracking. A system overview of the Lost Cosmonaut architecture is given in Fig. 7. It is based on a client-server infrastructure with iServer and OMSwe, a publishing framework, on the server side. The two server components manage all relevant information, whereas the client-side components provide the user interface.

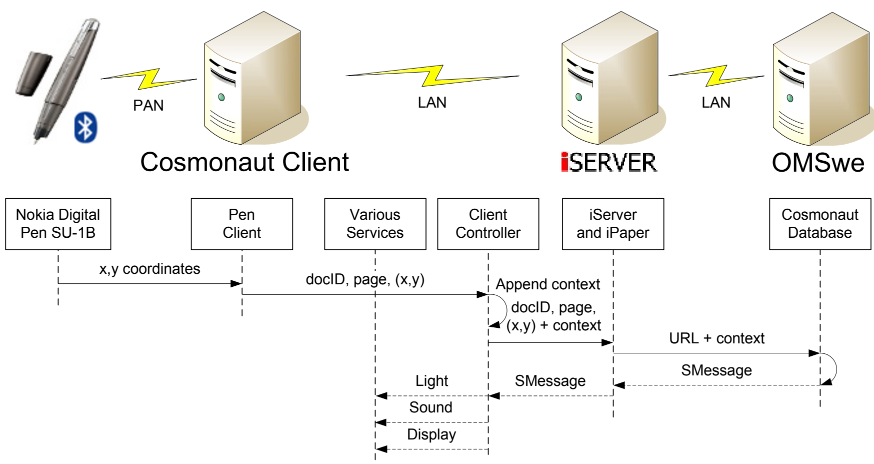


Fig. 7. Lost Cosmonaut architecture

As the user writes or points somewhere in one of the three documents with the Digital Pen, the `Pen Client` sends a request to iServer. The iPaper plug-in resolves the positional information and retrieves the appropriate object that is linked to the selected position. However, all application-specific information is stored in a separate OMSwe database and iServer only handles references to these applications objects in the form of OMSwe active components. The OMSwe database contains pre-authored information or data that has been added by previous visitors and returns the result as an XML Service Message (`SMessage`).

All requests from the client device to iServer are redirected over a `Client Controller` component. The `Client Controller` is responsible for dispatching the Service Message returned from iServer to the appropriate output channel. An important detail of the Lost Cosmonaut installation is that a Service Message may address multiple output channels. For example, if a new document is placed on the table, a request is sent to iServer and a Service Message containing information for all mood channels, i.e. the light control, the ambient sound system and the video display service, is returned. The architecture is flexible in supporting new output channels since only a new Service Message has to be defined and registered with the `Client Controller`.

8 Observations

We set up a demonstrator of the installation at the ETH Zurich for a week as a showcase of our work-in-progress from which we gained the following insights.

Pen handling: The usage of the pen as an interactive reading tool as such was not so easy to comprehend without introduction. We are considering to start each user session with a playful theme related introduction, including a hands-on guide to the usage of the pen itself.

Feedback and Reward: The audience was not always aware of the feedback for writing interaction such as the speeding up of the heartbeat. The reason is probably an inconsistency in how feedback is given. Consistent feedback appears crucial though as the user needs reassurance. Also the system only gives feedback when the user points at areas that contain some kind of visual content. Pointing at other areas doesn't trigger any reactions. We are considering some basic sound feedback for empty areas as a general reassurance.

Interaction with the Content: Some users tried out unexpected interaction patterns which proves disruptive at times. We will have to look closer at differing usage patterns in order to understand how they influence the experience.

Many users received the outputs of the system as rather randomly. This will be worth looking at because the more the user understands the consequences of his actions, the more likely he will engage himself with the installation

Narrative: This installation is setting up an environment with a specific theme, which emotionally involves and stimulates the user in order to fill the gaps with his own imagination. This works very well in some parts but the randomness of some of the output seems to be a bit disruptive in regards to the narrative.

9 Conclusion

The Lost Cosmonaut shows that digitally enhanced paper can offer a very interesting approach for developing interactive and nonlinear narrative. It opens up many possibilities for providing rich experience and involvement and it also supports imagination as well as user creativity. Our project combines the positive aspects of paper and pen providing a tactile embodied experience and relates those properties to a complex multimedia environment.

As to the narrative aspects of this installation: According to Kuitenbrower there are three main categories of interactive relations between user and system [3]: A "functional relation" where the system does exactly what it is told. On the contrary side of this experience there is "no relation" where the system seems to take no notice of the user but has its own life. Between these two extremes there is an "interesting relation" where "the system reacts to us with some kind of flavor of its own". This middle ground is where we position the Lost Cosmonaut. We will work further on providing more consistent feedback while sustaining an element of mystery and imagination. We are aware that the narrative structure we use, hardly supports a classical dramatic progression nor does it have dramatic closure. But this is the price that non-linear storytelling has to pay for exploring new areas. However we can easily imagine how the infrastructure could be used to provide a more linear or even game-like experience.

We would like to thank Andrea Lioy for his work on conceptualising and writing, Kurt Schlegel for sound composing and engineering as well as Rudi Belotti, Corsin Decurtins, Ljiljana Vukelja, Nadir Weibel for vital contributions.

References

1. M. Back, J. Cohen, R. Gold, S. Harrison, and S. Minneman. Listen Reader: An Electronically Augmented Paper-Based Book. In *Proceedings of ACM CHI 2001, Conference on Human Factors in Computing Systems*, Seattle, USA, March 2001.
2. J. J. Gibson. *The Senses Considered as Perceptual Systems.* Houghton Mifflin, Boston, 1966.
3. K. Kuitenbrower, Emergent Narrative, www.mediamatic.net/article-200.9160.html.
4. H. R. Maturana and F. J. Varela. *Der Baum der Erkenntnis – Die biologischen Wurzeln des menschlichen Erkennens.* Goldmann-Verlag, 1987.
5. A. Mazalek, G. Davenport, and H. Ishii. Tangible Viewpoints: A Physical Approach to Multimedia Stories. In *Proc. of 10th ACM Intl. Conference on Multimedia*, Juan-les-Pins, France, December 2002.
6. M. C. Norrie and B. Signer. Overlaying Paper Maps with Digital Information Services for Tourists. In *Proc. of ENTER 2005, 12th Intl. Conference on Information Technology and Travel & Tourism*, Innsbruck, Austria, January 2005.
7. A. J. Sellen and R. Harper. *The Myth of the Paperless Office.* MIT Press, 2001.
8. B. Signer and M. C. Norrie. A Framework for Cross-media Information Management. In *Proc. of EuroIMSA 2005, Intl. Conference on Internet and Multimedia Systems and Applications*, Grindelwald, Switzerland, February 2005.
9. P. Wellner. The DigitalDesk Calculator: Tangible Manipulation on a Desk Top Display. In *Proc. of ACM UIST'91, 4th Annual Symposium on User Interface Software and Technology*, November 1991.

Dream of Mee-Luck: Aspiration for a New Dawn*

Youngho Lee, Dahee Kim, Youngil Lim, Kyuhyung Kim,
Haesun Kim, and Woontack Woo

GIST U-VR Lab.,
Gwangju 500-712, S.Korea
{ylee, dkim, ylim, kh-kim, kim228, wwoo}@gist.ac.kr
http://uvr.gist.ac.kr

Abstract. Many countries in the world are putting much effort for the preservation of cultural assets of their own countries using digital. However, most of the attempts are based on the aspect of preservation science which aims to exquisitely restore cultural assets through historical investigation. Therefore in most cases they are successful in the restoration of shape without incorporating the spirit and meaning of the respective cultural asset. In this regard, we suggest an approach using Responsive Multimedia System which can present legend of Unju Temple as a sequence of events in VE. Users can experience virtual story by manipulating interactive StoryMap displayed on ARTable. In the suggested artwork, we express the uncompleted beauty, which we cannot find out from Unju Temple today.

1 Introduction

Many countries in the world are putting much effort for the preservation of cultural assets of their own countries. The development of a system for permanent preservation of cultural assets using digital restoration technology is one of them. Cultural assets carry their important values not only from the aspect of tangible shape as an object but also from the aspect of the spirit and legend related to the cultural asset. However, existing system has focused on the restoration of their shape. The suggested artwork, Dream of Mee-luck: Aspiration for a New Dawn, will re-illuminate true meaning of the cultural assets as it dealt with the restoration of cultural assets not only from physical aspect but also from spiritual aspect which did not exist so far in the restoration effort.

2 Artistic Motivation

Unju Temple is a temple located in Daecho-ri, Hwasun-gun, Jeonnam Province in Republic of Korea. In this area, a lot of stone pagodas and Buddhist statues are scattered in a group throughout the mountain and valley [1].

The legend has that around the end of Unified Shilla Dynasty Reverend Priest Doseon erected a thousand Buddhist statues and thousand stone pagodas only in one

* This work was supported in part by CTRC, and in part by MIC through RBRC at GIST.

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 280 – 283, 2005.

night. But there are no records of history about the constructions and ruins of those relics. People can just imagine as to when the thousand Buddhist statues and thousand stone pagodas were constructed, who constructed them and the reason why. However, there is no clear answer at all. Based on the relics found and excavated from the temple site, it is estimated that the age of the creation of Unju Temple would be around the ending of Silla in 11^{th} or 12^{th} century and the erections of statues and stone pagodas would have been done not in a day but over a long period of time.

Currently there are only 17 stone pagodas and 80 Buddhist statues left but we can find the vestiges of the thousand Buddhist statues and thousand stone pagodas in many places around the area. We call the Buddhist statues in Unju Temple as 'Mee-luck' (Maitreya: future buddha), Buddha who is to come to this world in the future and are our hope and wish toward the future. However, in Unju Temple now, Buddhist statues and stone pagodas exist but 'Mee-Luck' does not exist any more. Unju Temple has experienced the creations and extinctions through several reconstructions over a long period of a thousand years and overcome the pains of existence as well as non-existence against the flow of history. The shape of Unju Temple is now deformed and is in the pain of non-existence due to human greed.

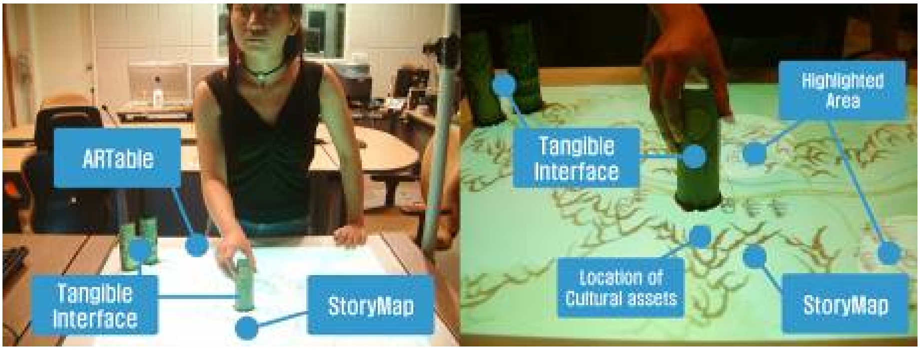
Consequently, we are left with vacant heart since it only exists in our heart. As Reverend Priest Doseon who dreamed to erect the thousand Buddhist statues and thousand stone pagodas, we are going to stand up the thousand Buddhist statues and thousand stone pagodas as well as lying-down Buddha in a new space called 'U-VR' (Ubiquitous Computing and Virtual Reality) aiming at building up the relationships of all global people where the hopes and wishes can be realized.

3 System Installation and Scenario

What we want to show in the suggested artwork are the wishes of ancient people which are incorporated in the thousand Buddhist statues and thousand stone pagodas as well as the unfinished beauty. To incorporate such meaning in the Artwork, we have materialized the following scenario: *Users watch animation or listen a legend of Unju Temple. They move to in front of ARTable and large display screen. They start to navigate virtual Unju temple. There are seven events to unfold the story as user's position in VE changes. In addition, users can experience Unju Temple repeatedly through 3D web site.*

We integrated ARTable, 3D Sound system, VRManager, and vr-UCAM into the RMS [2]. There were stereoscopic display system, a workstation for vision-based tracking system and one workstation for sound and database server. In addition to workstation, we installed two cameras and 5.1 channel sound system. When trigger signal is delivered to the sound server from client, it selects proper sound file to play through 5.1 channel speaker installed in the room. We built Database and Sound server to save 3D Model and to play 5.1 channel sound.

We implemented object and interactive StoryMap of Multi-modal Tangible User Interface to guide virtual story. We made dolls of a small priest, a tiger, and a dragon related to the story and participants can select one of them as his preference. For building stone pagodas, we made shovel-shape object. So, participants can move



(a) (b)

Fig. 1. (a) A user manipulate Tangible Interface, (b) Dolls (Tiger, Dragon, and Small Priest)

stone with shovel-shape object. Furthermore, we connected watch and computer to acquire current time. We used this watch to allow participants to change position of sun in VE

Interactive StoryMap which was displayed on ARTable of RMS is shown in Fig. 1. It was designed for guiding or causing user's interaction. The StoryMap showed geometric information of 'Unju Temple'. When participants invoked interaction, specific position lighted up to guide next position to go. We designed several StoryMap to select according to user's context. We also implemented Virtual Plant, Priest Doseon, and ruining Maaeyeorae in virtual environment. We created key frames of virtual object and combined them as a sequence by using pfSequence. A pfSequence is a pfGroup that sequences through a range of its children, drawing each child for a specified duration. Each child in a sequence was thought of as a frame in an animation. We controlled whether an entire sequence repeats from start to end, repeats from end to start, or terminates.

Fig. 2. Implemented Storytelling System. Left figure shows stonecutter on sunny day and right figure shows them on foggy day.

The user can interact with virtual environment through ARTable. As shown in Fig. 2, a user was able to navigate virtual environment manipulating objects on the ARTable. For example, the user was able to move a doll which was connected to virtual avatar in virtual environment. The user also tried to give special effect in virtual environment by moving object on the ARTable. The intelligent response of virtual plant was shown in large display screen. According to user's context such as gesture, preference, the virtual plant displayed various responses. The user can watch the changes of virtual surrounding according to change of real surrounding. Since it was difficult for us to acquire changes of weather or temperature, we simulated those changes by programming. We made watch which is connected to time of day in VE, so that when the user changed the watch, the time of virtual environment was changed. This demonstration showed us that when real and virtual environment are connected by context, it could improve immersion of user.

4 Conclusion

Unju Temple from the physical aspect which we can visit and see with our eyes clearly exists in reality. However, it is no longer possible to experience the spirit which was incorporated in the cultural asset now, unless we really try to find out, because of damage and new arrangement works on the cultural assets. Artwork 'Dream of Mee-luck' could be a starting point which will make us think about the restoration of cultural assets in true sense once again by providing the opportunity to experience various interactions and events on not only the physical aspect but also the spiritual aspect which has not existed so far.

References

1 Yochen Hiltmann, "Miruk: Die Heiligen Steine Koreas," *Edition Qumran im Campus Verlag*, 1993.
2 Youngho Lee, Sejin Oh, Beom-chan Lee, Jeung-Chul Park, Youngmin Park, Yoo Rhee Oh, Seokhee Lee, Han Oh, Jeha Ryu, Kwan H. Lee, Hong Kook Kim, Yong-Gu Lee, JongWon Kim, Yo-Sung Ho and Woontack Woo, "Responsive Multimedia System for Context-based Storytelling," *Proc. of Pacific-Rim Conference on Multimedia*, 2005 .
3 Youngho Lee, Sejin Oh, Woontack Woo, "Context-based Storytelling with Responsive Multimedia System," *Internal Conference on Virtual Storytelling,* 2005. (accepted)
4 Sj.Oh, Y.Lee, W.Woo, "vr-UCAM2.0: A Unified Context-aware Application Model for Virtual Environments," *ubiCNS*, 2005.

Interactivity and Digital Environments: Designing a Storymap for *Gormenghast Explore*

Julia Sussner and Maureen Thomas

Digital Studios, Department of Architecture,
University of Cambridge, UK
jms214@cam.ac.uk

Abstract. This paper discusses the process of conceptualising and graphically mapping narrative into a 3D computer generated environment using the example of the experimental interactive drama production of *Gormenghast Explore*. This mapping is represented by a Storymap, which combines spatially organised and reconfigurable narrative. Combining architecture theory and practice with interactive design, this new approach illustrates how principles of urban planning and character development work together to create a new template for the digital landscape.

1 Introduction

According to Marie-Laure Ryan, there are three types or immersion: "spatial immersion, the response to setting; temporal immersion, the response to plot; emotional immersion, the response to characters"[1] (Ryan, 2001: 121). This Storymap demonstrates how to incorporate these various points of immersion into the 3D digital environment.

1.1 Reconfigurable Narratives

As part of the EU Framework 6 IST Integrated Research Project, "NM2: New Media for the New Millennium"[2] the Cambridge University Moving Image Studio, CUMIS, is producing *Gormenghast Explore,* to guide and test the design of digital tools for handling interactive narrative via broadband. The aim is to re-edit the BBC millennium series *Gormenghast* (Wilson, UK 2000), televised January 2001, (4 hours of footage) into an interactive movie, which can be viewed in meaningful narrative segments of any length, configured in real time by the viewer, as a model for producing new, spatially organised interactive drama. The research project is a process of analysis and exploration though various narrative zones and pathways, to establish new frameworks for an architecturally organised fictive experience which avoids branching narrative and obtrusive interactions.

---

[1] Ryan, Marie-Laure (2001: 121).

[2] Williams, D.L, Kegel, I.C, Mayer, H, Stollenmayer, P. & Ursu, M., 2005; http://www.ist-nm2.org; 'NM2 New Media for a New Millennium' EWIMT Proceedings Nov 2004.

G. Subsol (Ed.): VS 2005, LNCS 3805, pp. 284 – 287, 2005.

The opportunity to explore the narrative through selected parameters enables the interactor to understand complex relationships that would not otherwise be apparent in a pre-determined linear non-interactive narrative. The BBC footage is based on Mervyn Peake's *Gormenghast Trilogy*. Through the three novels, *Titus Groan* (1946), *Gormenghast* (1950), and *Titus Alone* (1959), Peake constructs a gothic narrative around the castle of Gormenghast. Partly as an allegory of the British class system

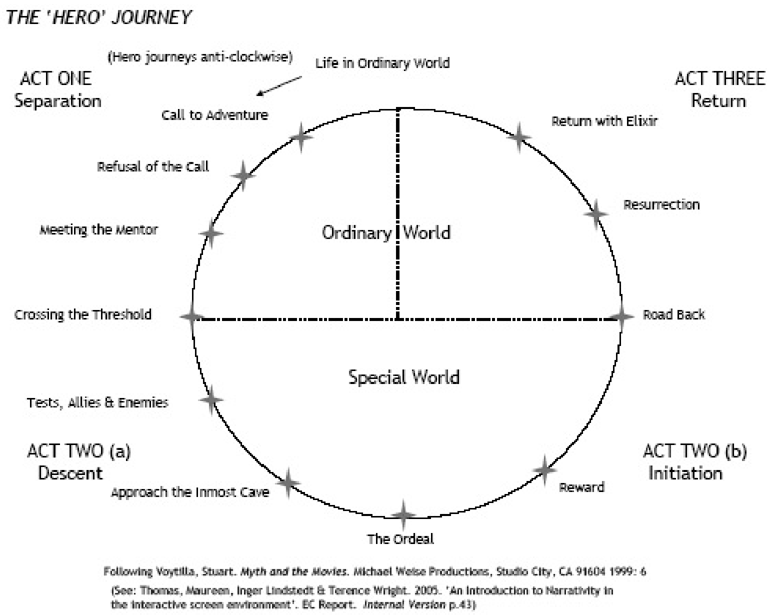


Fig. 1. Epic Heroic Journey

and partly a tale of the decline of Europe, Peake weaves a complex episodic narrative around the castle and the cast of characters who inhabit it. The episodic and spatial qualities of the novels make them extremely well suited for reconfigurable stories. Multiple versions of the 'hero tale'[3] and plots such as the 'Romance', the 'Revenge Play' and the 'Contest'[4], diversify the potential of the material to form compelling narratives. This expansion requires an infrastructure to guide the interactor. The user interface must be designed so that the fashioning of space becomes a pattern language[5] to understanding the content. Figure 1 shows the Epic Heroic Journey, which is also integrated into the Storymap (Figure 2).

3 Voytilla, Stuart (1999).

4 Thomas, Maureen, Inger Lindstedt & Terence Wright. 'An Introduction to Narrativity in the interactive screen environment'. EC Report. www.ist-nm2.org (2005).

5 Alexander, Christopher (1977).

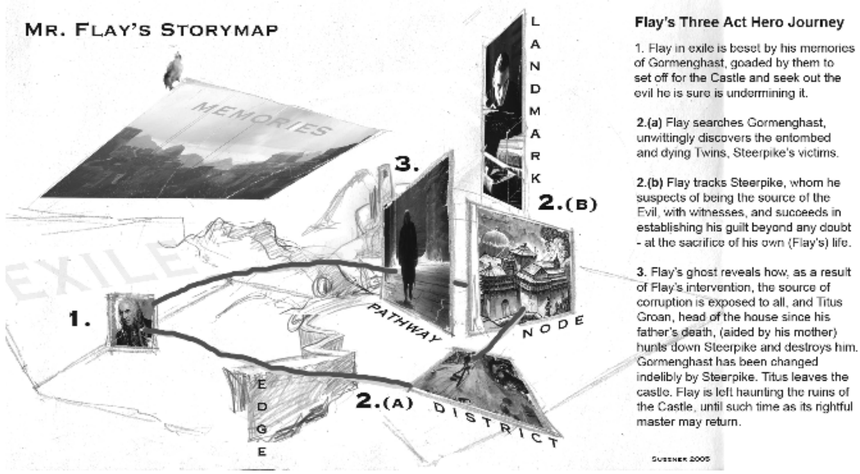


Fig. 2. *Gormenghast Explore* Storymap for the character of Flay. The multimedia clips are spatially-organised so that the inherent nature of the planning informs the interactor. The Epic Heroic Journey is an inhabitable space.

2 Castle as City

Gormenghast Explore exploits the organisation of the story into scenes played out in the different zones of the Castle of Gormenghast, each of which is characterised - in the original novels and the production-design of the TV serial - by marked emotional features. Few of the inhabitants of the Castle are free to roam all the nooks and crannies it offers; some have pathways fixed by habit and status, and there are clearly defined edges to the districts they journey through between the landmarks of the monumental, city-like storyscape of Gormenghast. The narrative pathways of characters and plots intersect at the vital nodes of the Castle's expressive space.

Urban design theory considers the responsibility to provide spaces that generate meaning through design for the inhabitant. In *The Image of the City,* Lynch describes the process and necessities of creating sites of interaction: "although clarity or legibility is by no means the only important property to a beautiful city, it is of special importance when considering environments at the urban scale of size, time and complexity. To understand this, we must consider not just the city as a thing in itself, but the city being perceived by its inhabitants"[6]. It is in this context that he parses the urban environment into the five categories: nodes, pathways, districts, edges, and landmarks, which are instantiated in the analysis and synthesis of the *Gormenghast Explore* Storymap. 'The observer himself should play an active role in perceiving the world and have a creative part in developing its image. He should have the power to change that image to fit changing needs,'[7]. In this experimental production, *Gormenghast Explore*, interactors are perceived as such active observers, exploring and

[6] Lynch, Kevin (1960: 3).
[7] Lynch, Kevin (1960: 6).

reconfiguring a personal narrative cityscape– like Flay, gradually mapping their own version of the Castle of Gormenghast and the tales of its inhabitants.

The Situationist theory group (notably Guy Debord) explored the process of re-thinking the map as a subjective story-making legend in the 1950's. Interactive cinema shares with the Situationists the desire to rethink expressive, inhabited space according to subjective point of view, by creating an environment based on spatialised actions or narratives. Indeed, for both, "space becomes a process"[8]. Yet where the Situationists remain within the context of the two-dimensional street map, *Gormenghast Explore* employs 3D technology and architectural thinking to create a legible and coherent Storymap which patterns the personalized journey of the interactor as an architectural program of use. Architectural design and theory as applied to 3D computer generated environments is a constructive and fruitful tool to construct the dwelling of spatially organised narrative and interactive cinema[9].

Integrating architectural theory and the heroic journey into interactive design, the Storymap synthesises Ryan's three points of immersion (spatial→setting, temporal→plot, and emotional→characters). The Storymap creates a mise-en-scène to inform and support the interaction and ultimately, the performance.

References

1. Alexander, Christopher: A Pattern Language: Towns, Buildings, Construction. Oxford University Press, New York (1977)
2. McDonough, Thomas: Situationist Space. October, n.67, winter (1994)
3. Ryan, Marie-Laure: Narrative as Virtual Reality: Immersion and Interactivity in Literature and Electronic Media. Johns Hopkins University Press, Baltimore (2001)
4. Thomas, Maureen: Playing with Chance and Choice – Orality, Narrativity and Cinematic Media. in Bushoff, Brunhild ed. Developing Interactive Narrative Content: sagas/sagas.net reader. High Text, Munich (2005) 371-442
5. Thomas, Maureen and Francois Penz (eds.): Architectures of Illusion: From Motions Pictures to Interactive Navigable Environments. Intellect Books, Bristol (2003)
6. Thomas, Maureen, Inger Lindstedt & Terrance Wright: Plot, Story, Screen: An Introduction to Narrativity. EC Report NM2 D2.1 – Public Deliverable. www.ist-nm2.org
7. Voytilla Stuart: Myth and the Movies: Discovering the Mythic Structure of 50 Unforgettable Films. Michael Weise Productions, Studio City (1999)

[8] McDonough, Thomas (1994: 68).
[9] For more on spatially organised narrative see Thomas, 2003 and 2005.

Author Index