LNCS 6243

Hyun Seung Yang
Rainer Malaka
Junichi Hoshino
Jung Hyun Han (Eds.)

Entertainment Computing – ICEC 2010

9th International Conference, ICEC 2010
Seoul, Korea, September 2010
Proceedings

ifip

Springer

Lecture Notes in Computer Science 6243

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Hyun Seung Yang Rainer Malaka
Junichi Hoshino Jung Hyun Han (Eds.)

Entertainment Computing - ICEC 2010

9th International Conference, ICEC 2010
Seoul, Korea, September 8-11, 2010
Proceedings

Springer

Volume Editors

Hyun Seung Yang
KAIST, Department of Computer Science
373-1 Kusung-dong Yusung-gu, Daejeon, Korea 305-701
E-mail: hsyang@kaist.edu

Rainer Malaka
Universität Bremen, Technologie-Zentrum Informatik (TZI)
Postfach 330440, 28334 Bremen, Germany
E-mail: malaka@tzi.de

Junichi Hoshino
University of Tsukuba
Graduate School of Systems and Information Engineering
1-1-1 Tennoudai, Tsukuba, Ibaraki, 305-8573 Japan
E-mail: jhoshino@esys.tsukuba.ac.jp

Jung Hyun Han
Korea University, Department of Computer Science and Engineering
Anam-dong Seongbuk-gu, Seoul, Korea 136-701
E-mail: jhan@korea.ac.kr

Library of Congress Control Number: 2010932769

CR Subject Classification (1998): H.5.2, I.3, I.5, I.6, I.4, I.2

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN 0302-9743
ISBN-10 3-642-15398-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-15398-3 Springer Berlin Heidelberg New York

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

The 9th International Conference on Entertainment Computing (ICEC 2010) was held in September 2010 in Seoul Korea. After Pittsburgh (2008) and Paris (2009), the event returned to Asia.

The conference venue was the COEX Exhibition Hall in one of the most vivid and largest cities of the world. This amazing mega-city was a perfect location for the conference. Seoul is on the one hand a metropolitan area with modern industries, universities and great economic power. On the other hand, it is also a place with a very fascinating historical and cultural background. It bridges the past and the future as well as east and west.

Entertainment computing also aims at building bridges from technology to leisure, education, culture and work. Entertainment computing at its core has a strong focus on computer games. However, it is not only about computer games. The last ICEC conferences have shown that entertainment computing is a much wider field. For instance in games, technology developed for games can be used for a wide range of applications such as therapy or education. Moreover, entertainment does not necessarily have to be understood as games. Entertainment computing finds its way to stage performances and all sorts of new interactive installations.

In this sense, entertainment computing has the potential of influencing many other areas of computer science and engineering. But in contrast to other disciplines, entertainment computing always starts from the human experience and with the question of how to tell the users a story and how to immerse people into an entertaining world. Many critics of modern computer games believe that these technologies may distract people from the real world. In contrast to this, however, we are exploring the new medium of digital entertainment in order to expand classic media and their expressive power.

The presentations and papers of ICEC 2010 showed the broad spectrum of entertainment computing. The paper selection process was selective. Out of over 100 papers, 19 were accepted as long papers and 27 as short papers. With poster presentations we give researches the opportunities to present new ideas and also work in progress. The review process involved an international board of reviewers who put much effort into selecting an excellent set of papers for presentation.

For the highly influential conference, in addition to paper presentations, we put utmost effort in bringing together a number of distinguished keynote and invited speakers, global leaders representing a broad spectrum of entertainment computing fields; Roy Ascott (Media Theory, Art& Science), Susumu Tachi (VR&Telepresence), Keith Devlin(Media & Human Science), Don Marinelli (Entertainment Technology), George Joblove (Movie Industry), Shigeru Saito (Game Industry), Naoko Tosa (Culture &Media Art), Norico Wada (Animation&Contents), Junichi Osada (HRI&

Robot Design), Zenjiro (Comedy&Performance), Tomonaka Takahashi (Robot Design). We also organized five workshops dealing with hot issues including Culture Computing and Media Arts, 3D Stereoscopic Technology and Contents Production, The Awakening of Asian Animation and Content Power, Improvisational Acting, Spatial Sound and Entertainment Computing.

We would like to thank all Organizing Committee/Program Committee members and all supporting organizations that helped to make this event possible: The IFIP as the leading supporting organization of ICEC and the IFIP TC14 as the sponsoring organization. We specially thank Yeong Nam Chae and all other graduate students and staff of the KAIST AIM Lab for their tremendous effort in the preparation and assistance of all main programs and wonderful activities of ICEC2010.

July 2010

Hyun Seung Yang
Rainer Malaka,
Junichi Hoshino
Jung Hyun Han

Organization

Honorary Chairs

Roy Ascott	Planetary Collegium and CAiiA, USA
Susumu Tachi	University of Tokyo and Keio University, Japan

General Chairs

Hyun Seung Yang	KAIST, Korea
Ryohei Nakatsu	IDMI NUS, Singapore IFIP TC14 Chair
Don Marinelli	ETC CMU, USA

Organizing Chairs

Matthias Rauterberg	Eindhoven University of Technology, The Netherlands
Fumio Kishino	Osaka University, Japan
Young Jik Lee	ETRI, Korea

Program Chairs

Junichi Hoshino	University of Tsukuba, Japan
Rainer Malaka	University of Bremen, TZI, Germany
Jung Hyun Han	Korea University, Korea

Media Art/Design Chairs

Naoko Tosa	Kyoto University, Japan
Oliver Grau	Danube University, Austria

Workshop Chairs

Yoomi Choi	Ewha University, Korea
Dongho Kim	Soongsil University, Korea

Exhibition Chair

Jiyoung Lee	CMU ETC, USA

Program Committee

Lynne Baillie (United Kingdom), Ville-Veikko Mattila (Finland), Christoh Bartneck (The Netherlands), Kazunori Miyata (Japan), Hisham Bizri (USA), Hiroshi Mori (Japan), Anne Braun (Germany), Noriko Nagata (Japan), Tony Brooks (Denmark), Masayuki Nakajima (Japan), Pedro Gonzalez Calero (Spain), Atsushi Nakano (Japan), Marcello Carrozzino (Italy), Ryohei Nakatsu (Singapore), Marc Cavazza (United Kingdom), Stephane Natkin (France), Adrian David Cheok (Singapore), Anton Nijholt (The Netherlands), Sung Hyun Cho (Korea), Seiichi Nishihara (Japan), Yoo Mi Choi (Korea), Kazushi Nishimoto (Japan), Hyung IL Choi (Korea), David Obdrzalek (Czech Republic), Paolo Ciancarini (Italy), Kyu Whan Oh (Korea), Philippe Codognet (Japan), Hiroshi Okuno (Japan), Michael Cohen (Japan), Elina M.I. Ollila (Finland), Fionnuala Conway (Ireland), Tetsuo Ono (Japan), Nuno Correia (Portugal), Zhigeng Pan (China), Sidney Fels (Canada), Lucia Pannese (Italy), Newton Fernando (Singapore), Chang Joon Park (Korea), Brian Fisher (Canada), Jisoo Park (Korea), Sam Ge (Singapore), Jong-II Park (Korea), Nahum Gershon (USA), Lyn Pemberton (United Kingdom), James Gimzewski (USA), Yusuf Pisan (Australia), Jung Hyun Han (Korea), Cristina Portales (Spain), Stephen Hansen (Australia), Peter Purgathofer (Austria), Brenda Harger (USA), Matthias Rauterberg (The Netherlands), Jaap van den Herik (The Netherlands), Richard Reilly (Ireland), Junichi Hoshino (Japan), Zhong Ren (China), Xian-Sheng Hua (China), Marco Roccetti (Italy), Michita Imai (Japan), Maria Roussou (Greece), Masa Inakage (Japan), Ben Salem (The Netherlands), Takehiko Kamae (Japan), Huaqin Shen (China), Haruhiro Katayose (Japan), Akihiko Shirai (Japan), Hirokazu Katoh (Japan), Kazuto Shiratori (Japan), Gerald J Kim (Korea), Andy Sloane (Unite Kingdom), Laehyun Kim (Korea), Masanori Sugimoto (Japan), Dong Ho Kim (Korea), Yasuyuki Sumi (Japan), Yoshifumi Kitamura (Japan), Kaoru Sumi (Japan), Fusako Kusunoki (Japan), Barnabas Takacs (Hungary), Jong-Weon Lee (Korea), Hiroki Takahashi (Japan), Inkwon Lee (Korea), Tsutomu Terada (Japan), Joon Suk Lee (Korea), Masashi Toda (Japan), Chil Woo Lee (Korea), Naoko Tosa (Japan), Seungyon Lee (Korea), Sofia Tsekeridou (Greece), Ji Young Lee (USA), Lucia Vera (Spain), Soon Bum Lim (Korea), Roel Vertegaal (Canada), Duh Been Lim (Singapore), Victoria Vesna (USA), Artur Lugmayr (Finland), Yangsheng Wang (China), Luzhuang Ma (China), Kevin Wong (Australia), Rainer Malaka (Germany), Woontack Woo (Korea), Tony Manninen (Finland), Yasuyuki Yanagida (Japan), Donald Marinelli (USA), Xubo Yang (China), Panos Markopoulos (The Netherlands), Hyun Seung Yang (Korea), Marco A. Gomez Martin (Spain), Sung Eui Yoon (Korea), Javier Jaen Martinez (Spain), Radslav Yoshinov (Bulgaria), So Masuko (Japan), Ruwei Yun (china), Hitoshi Matsubara (Japan), Alexander Markowetz (Germany), Tim Marsh (Singapore)

Supporting Organizations

International

International Federation for Information Processing (IFIP)
Institute of Electrical and Electronics Engineers (IEEE)

Association for Computing Machinery (ACM)
Information Processing Society of Japan (IPS)
The British Computer Society (BCS)
Societe des electriciens et des electroniciens (SEE)
Gesellschaft für Informatik e.V. (GI)
Asia Digital Art and Design Association (ADADA)

Local

The Korean Institute of Information Scientists and Engineers (KIISE)
Korea Multimedia Society (KMMS)
Human Computer Interaction KOREA (HCI KOREA)
Korea Computer Graphics Society (KCGS)
Korea Game Society (KGS)
Electronics and Telecommunications Research Institute (ETRI)
KAIST Microsoft Research Collaboration Center (KMCC)

Table of Contents

Long Papers

New Interfaces and Entertainment Robots

User Interfaces

Serious Games and Collaborative Interaction

Tools and Networks

Game Theory and User Studies

Short Papers

Game Theory, User Studies and Story Telling

Interaction and User Experience

Serious Games

Tools and Methods

Robots and New Interfaces

Posters

Baby Robot "YOTARO"

Hiroki Kunimura[1], Chiyoko Ono[1], Madoka Hirai[2], Masatada Muramoto[1], Wagner Tetsuya Matsuzaki[1], Toshiaki Uchiyama[1], Kazuhito Shiratori[3], and Junichi Hoshino[3]

[1] University of Tsukuba, Graduate School of Comprehensive Human Sciences
[2] University of Tsukuba, Graduate School of Life and Environmental Sciences
[3] University of Tsukuba, Graduate School of Systems and Information Engineering
305-8577 Tennodai 1-1-1, Tsukuba, Japan
kunimura@geijutsu.tsukuba.ac.jp, onochiyoko@geijutsu.tsukuba.ac.jp, M.hirai@aist.go.jp, wagner@geijutsu.tsukuba.ac.jp, muramasa@geijutsu.tsukuba.ac.jp, uchi@kansei.tsukuba.ac.jp, c_tori@mbb.nifty.ne.jp, jhoshino@esys.tsukuba.ac.jp

Abstract. YOTARO is a baby-type robot developed to create a new communication perspective between robots and humans through interaction experience based on the reproduction of a baby's behaviors and user actions. YOTARO exhibits different emotions and reactions, such as smiling, crying, sleeping, sneezing, and expressing anger. It is controlled by an emotion control program that executes in response to inputs such as touching its soft and warm face, touching its stomach, and shaking a rattle. The output is in the form of interactive reactions such as emission of sounds, change of expressions, limb movements, sniveling, and variation in skin color. In addition, we used questionnaires to observe the impression on users before and after their experience with YOTARO.

Keywords: interaction, communication, virtual reality, physical contact.

1 Introduction

The relationship between humans and robots, as well as the communication between them, has garnered increasing attention as robots and virtual reality (VR) systems will play an important role in family lives in the future.

Previous research on this subject, Infranoid [1] and Keepon [2], tried to clarify human communication. Kismet [3] is an example of emotional reactions and turn-taking conversations in robots. Another case refers to a media artwork, Neuro-Baby [4] with emotional reactions. There are also robot types whose main objective is creating long-term relationships with people, such as the mental-commit robot Paro [5], AIBO [6] the pet-type robot, and mobile games such as Nintendos [7] and Tamagotchi [8].

This research aims to create a new communication perspective through a baby-type robot, YOTARO. The kindness toward babies is inherent in humans and the baby is the most conspicuous example of human instincts, in which everyone cares for and is tender toward the baby.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 1–8, 2010.

The actions of caring for a baby, such as to lull a baby or to wipe a runny nose create a strong connection with the baby. A similar connection can be established in the relationship between humans and robots. YOTARO aims to create a feeling of satisfaction in people through only a few minutes of interaction with the baby. Fig. 1 shows the physical appearance of YOTARO.

Fig. 1. Physical appearance of baby-type robot “YOTARO”

2 Experience

YOTARO has six elements that correspond to the interactions with a baby.

- Baby’s peculiar transient behavior
- Unconsciously touching its ruddy and soft skin
- Slightly higher body temperature (peculiarity of babies)
- Help to wipe snot
- Rattle as an example of communication using tools
- Movements that stimulate touching a partner.

The experience booth is the interior of a child’s room and includes a baby bed; the users stand beside the bed and play with YOTARO.

The user starts by waking up the sleeping YOTARO merely by touching its face. Next, the user can perform actions such as touching its stomach, shaking a rattle, and touching its face. YOTARO reacts by changing its expressions and skin color, sniveling, sneezing, emitting sounds, and moving its limbs. In this process, the user performs various actions to make YOTARO happy, and the pseudo experience of taking care of a baby is improved when the user wipes the snot.

YOTARO wakes up in a bad humor and its reaction is inconsistent. However, its humor gradually gets better and laughter is heard. While users are enjoying this experience, the reactions slowly get worse. YOTARO does not react pleasantly to actions as it had done until now, and it is now almost crying. This situation occurs when a baby is getting tired and being fretful. Based on this condition, the users can understand how to take care of a baby as every action now is likely to make the baby more fretful. In this case, after a short while of doing nothing, YOTARO sleeps again. The experience ends with completion of this cycle. Fig. 2 shows a simplified image of YOTARO's structure.

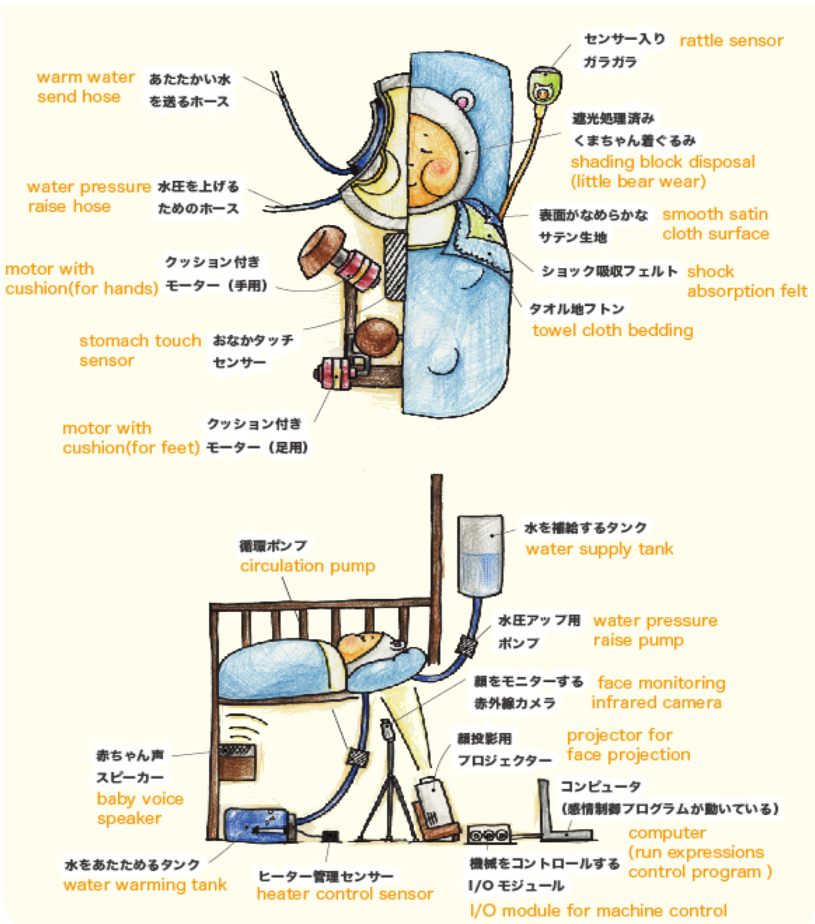


Fig. 2. Structure of YOTARO

3 System

3.1 Summary of System

The YOTARO system is presented to better understand the general structure and its interactive reaction. The system is divided into three parts: input, controller, and output.

The user action input is captured through various sensor devices; the controller transforms the input data into various reactions, and the system then produces output actions such as sounds, tactile sense, snivelling, and movements to interact with users (Fig. 3).

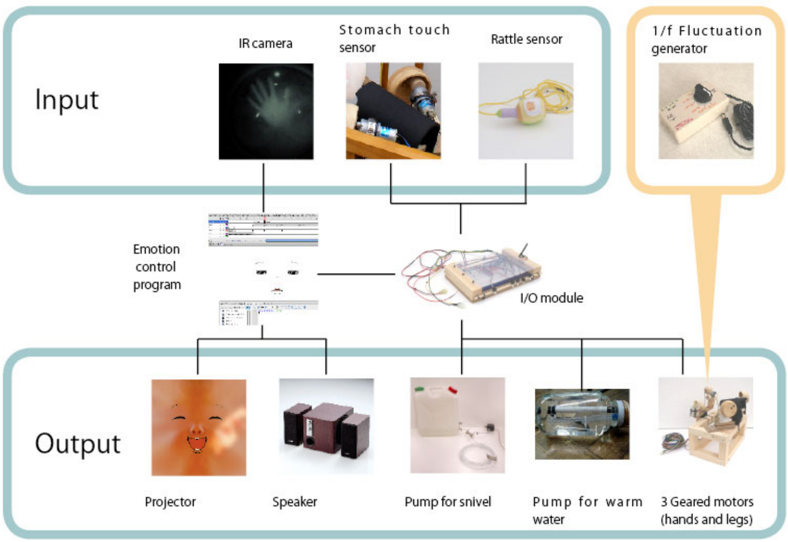


Fig. 3. Relationship between system configuration and Input/Output

3.1.1 Input

The virtual experience of contacting a baby happens when users touch YOTARO's face. It is necessary to distinguish which part of the face was touched. Three devices compose the inputs: near-infrared camera, touch sensors, and rattle sensor.

First, the near-infrared camera is installed on the backside of YOTARO's translucent face; it detects the position of the hand by infrared ray reflection. Next, touch sensors detect actions such as tickling the stomach using a photodiode switch. For rattle, piezoelectric elements are sewed and connected via a cable. Upon shaking the rattle, values of the sensor change and this change is detected via an analog recorder.

3.1.2 Controller

These three kinds of inputs are controlled through the I/O module and an emotion control program that reads the input, processes the data, and outputs the reactions. The emotion control program controls the changes of expressions; for example, when its nose is touched, YOTARO sneezes; when its cheeks are touched, it becomes happy. Thus, the baby promptly changes emotions and the output responses from the sleeping state, to waking up, being happy, being fretful, and then sleeping again.

3.1.3 Output

The outputs are produced in real time through the following types of methods: a video projector that plays an animated movie, audio speakers, geared motors, a snivel pump, and a warm water pump. The video projector is used to project changes in the baby's expressions. The speakers emit a real baby's sounds that are sampled and emitted in connection with expressions. A group of three motors moves the hands and legs; there is one motor for each hand and one for both legs. The motors produce irregular movements by rotation under the blankets; a 1/f fluctuation is reflected on the movements. In most cases, the rhythms of the natural world have a 1/f fluctuation, such as in a small river stream or the chirping of insects. YOTARO moves its limbs as naturally as possible, and fluctuations were added to motor's rotation.

The pump for the snivel comprises a warm water pump and a tank. The nose area has small holes, and when the pump increases the pressure, drops of warm water start to flow. The water is warmed in the reserve tank by a thermostat and heater; the water circulation inside the face portion of the robot is performed by the water pump. The warm water is maintained inside the face at temperature that is near the normal human body temperature.

3.2 Emotion Control Program

YOTARO is a robot that can express emotions. In this section, previous research about robots that can display emotions was investigated as a base to develop the emotion control program for YOTARO.

Some cyclical aspects were noticed, such as the process of a sleeping baby waking up, the slow movements and reactions going into an awakened state, becoming tired after a while, sleepy, and in a bad mood, and finally sleeping again. This cycle was summarized as: sleep mode → doze mode → in a good mood mode → fretfulness mode → sleep mode, to represent it more closely to the emotion mode loop.

3.3 Methodology to Create Face for Projection

The baby's facial reactions are presented by animation, as shown in Fig. 4. Inside the face, the nose is the most difficult part to change expressions. Upon creating the baby's image, the nose is added. Over this, the emotion control program determines expressions and composes skin color changes. The near-infrared camera detects movements by capturing every frame and bitmap, and after that it compares them with the previous bitmap image. In instances where there are huge differences, skin becomes light yellow, while in small places, it becomes red. Through these effects, the pressured place creates the illusion of color change. Finally, the face composition is projected to form a final image at the translucent silicon face.

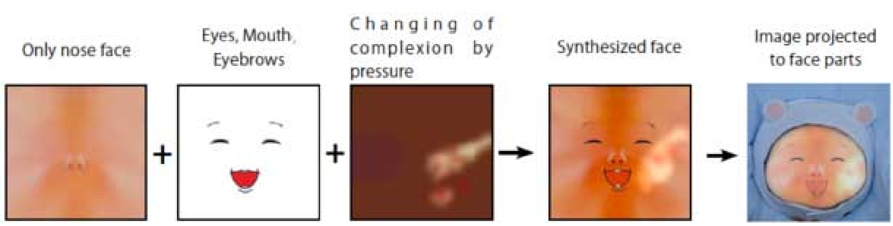


Fig. 4. Image formation of face

3.4 Structure of Face

The face portion of YOTARO has diverse functions: soak water, maintain soft and curve surface, and detection of the user's hand position. The face portion is a new type of back projection display. In addition, an approach was followed to maintain the user's interest while in the experience booth. These simple cycles when combined became more complex to keep the user interested in playing with YOTARO.

Fig. 5 shows the structure of the face. It was made using 2-mm-thick translucent silicon film, which covers a water system comprising two acrylic hemispheres.

The snivel flow control was realized by opening minute holes through the silicon film and acrylic surface. It works as a valve owing to silicon's elasticity, and changes internal pressure by on/off of the pump, thereby controlling the snivel flow.

The acrylic water system has four connectors for hoses. The connector on the top (Fig. 5-a) is to suck air, the one on the right (Fig. 5-b) is for the pump to send water from the tank, consequently elevating the pressure and causing the snivel to flow. The remaining two connectors (Fig. 5-c, d) are used to circulate warm water from the external source to the internal water system. This circulation permits the water temperature control inside the acrylic water system, and the surface temperature becomes similar to a baby's temperature.

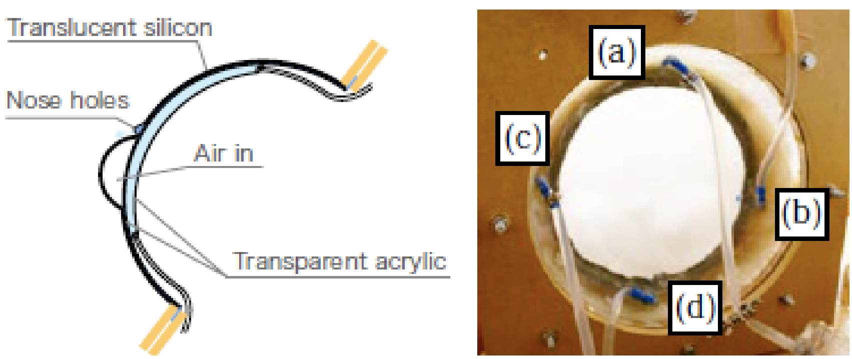


Fig. 5. Face structure

4 Variation and Considerations

During the exhibition of YOTARO at Laval Virtual 2009, data was collected from 75 people through a questionnaire (in English). The questions comprised five stages, at the first contact and after playing with YOTARO. The stages were as follows: "Very Good," "Good," "Whatever," "Bad," and "Very Bad." Moreover, the questionnaire inquired whether they have children.

The valuation of user's impression of YOTARO was presented in two averages: before and after play (Fig. 6). The average of user's impressions was compared between before experience and after experience cases, and it varied from 3.72 to 4.08.

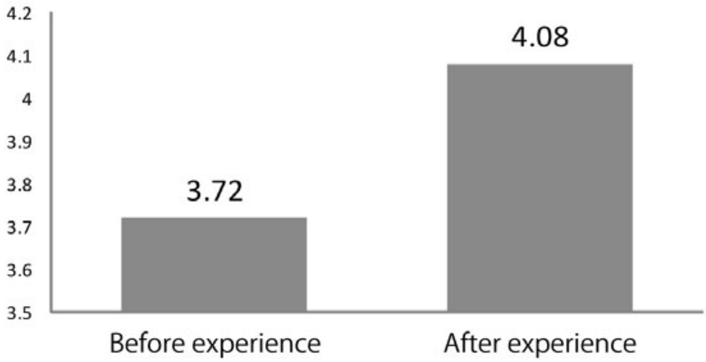


Fig. 6. Evaluation of impression of YOTARO

It was observed that the user's average of impression before experience had the tendency to rise than that of after the experience. Thus, it is possible to consider that the action of taking care of the baby robot significantly influences the affection of the doer..

5 Conclusion

This study was conducted on the hypothesis that the actions of taking care of babies create a strong connection with them. As an outcome of the materialization of an artificial baby, it became possible to develop a baby-type robot, YOTARO. Through the responses in questionnaires, it was observed that the impressions of users changed significantly before and after the contact with YOTARO. In addition, the hypothesis of "action of taking care" significantly influenced the affection with the "subject of this action" was supported through the results of the questionnaire data. Based on the presented arguments, to leave an impression more effectively from users to the robots, it is necessary to approach a situation where a robot needs to be taken care of and users are in a position to support this action.

References

1. Kozima, H., Yano, H.: A Robot that Learns to Communicate with Human Caregivers. In: Proceedings of the First International Workshop on Epigenetic Robotics (2001)
2. Kojima, H., Nakagawa, K.: Child-Robot Interaction in Therapeutic and Pedagogical Fields. Technical report of IEICE. HCS, Japan, 25–30 (2006)
3. Breazeal, C., Scassellati, B.: Infant-like Social Interactions between a Robot and a Human Caregiver. Adaptive Behavior 8(1), 49–74 (2000)
4. Kakimoto, M., Tosa, N., Mori, J., Sanada, A.: Neuro-Baby; The Automatic Facial Expression Synthesizer that Responds to Human Voice by Recognizing the Feelings. ITEJ Technical Report 16(33), 7–12 (1992)
5. Saito, M., Naruse, K., Kakazu, Y.: Analysis on Relation between AIBO's Motion and its Impression using Statistical Method. Technical report of IEICE. HIP, Japan 103(166), 23–28 (2003)

6. Wada, K., Shibata, T., Saito, T., Sakamoto, K., Tanie, K.: Long-term Robot Assisted Activity at a Health Service Facility for the Aged. The Japan Society of Mechanical Engineers, Japan (2005)
7. Nintendo (2005), `http://www.nintendo.co.jp/ds/adgj/`
8. Bandai (1996), `http://tamagotch.channel.or.jp/`
9. Miwa, H., Ito, K., Takanobu, H., Takanishi, A.: Development of a Human-like Head Robot for Emotional Communication with Human: 3rd Report. Introduction of the Equations of Emotion and Robot Personality, Transactions of the Japan Society of Mechanical Engineers C 70(699), 3244–3251 (2004)
10. Ekman, P.: An Argument for Basic Emotions. Cognition & Emotion 6, 169–200 (1992)
11. Kamiyama, K., Kajimoto, H., Kawakami, N., Tachi, S.: Evaluation of a Vision-based Tactile Sensor. In: IEEE International Conference on Robotics and Automation, vol. 2, pp. 1542–1547 (2004)
12. Kakehi, Y., Jo, K., Sato, K., Minamizawa, K., Nii, H., Kawakami, N., Naemura, T., Tachi, S.: ForceTile: Tabletop Tangible Interface with Vision-based Force Distribution Sensing. In: SIGGRAPH New Tech. Demos (2008)
13. Toshiki, S., Haruko, M., Hideki, K., Kentaro, F.: Photoelastic Touch: Transparent Rubbery Tangible Interface using an LCD and Photoelasticity, Photoelastic Touch. In: Proceedings of the 22nd Annual ACM Symposium on User Interface software and Technology (UIST 2009), Japan, pp. 43–50 (2009)

A Card Playing Humanoid for Understanding Socio-emotional Interaction

Min-Gyu Kim* and Kenji Suzuki

University of Tsukuba, Dept. of System Information Engineering,
Tennodai, 1-1-1, Tsukuba, Japan
mingyu@ai.iit.tsukuba.ac.jp, kenji@ieee.org
http://www.ai.iit.tsukuba.ac.jp

Abstract. This paper describes the groundwork for designing a social and emotional interaction between a human and robot in game-playing. We considered that understanding deception in terms of mind reading plays a key role in realistic interactions for social robots. In order to understand the human mind, the humanoid robot observes nonverbal deception cues through multimodal perception during poker playing which is one of human social activities. Additionally, the humanoid manipulates the real environment which includes not only the game but also people to create a feeling of interacting with life-like machine and drive affective responses in determining the reaction.

Keywords: Human-robot social interaction, Mind reading, Deception, Humanoid playmate.

1 Introduction

In successful social interactions, humans understand and manipulate other people's behavior in terms of their mind. These capabilities are related to development of Theory of Mind (ToM) in developmental psychology. ToM is the ability to understand others' internal states and the relationship between human behaviors. Children with ToM understand that the others have mental states such as belief, motivation, emotion and intentions, and the mental states often cause behaviors. ToM is developed into the ability to infer the others' thought, desire and emotion, and the ability to predict the others' behavior based on their inference. A normally-developed children in 4-5 ages can understand the fact that belief plays a leading part in behavior and can distinguish the self desires from the other's belief. Also, the children can understand that the other's behavior can be determined from the belief and the knowledge based on perceptual experiences.

In this study, we focused on deception which is an important part of human social competence. According to ToM, the skills of deception in complex form

* This work is partially supported by Global COE Program on "Cybernics: fusion of human, machine, and information systems."

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 9–19, 2010.

are developed from false belief[1] which is a social cognitive ability to know that beliefs can be false and that they can be manipulated[2]. In the autistic children's case, the impaired false belief comprehension affects their social interaction and communication. From this point of view, forthcoming social robots should be able to cope with complicated social skills of human with the embedded cognitive ability to infer concealed intentions and emotions from social contexts and predict consequent behaviors.

Some of the preceding researches have made effort to develop social robots to understand human mind in a wide scope. The purpose of Leonardo underlies socially guided learning and social referencing[3],[4]. The embodiment makes it more capable of forming emotional attachment with humans. KASPAR is being investigated for cognitive development research[5]. In particular, it shows the possible use of therapeutic or educational robotic systems to encourage social interaction skills in autism children. Keepon is designed to study social development by communicating with children[6]. Its behaviors are intended to help children to understand its attentive and emotive actions. Most of the social robot related researches attain to mimic cognitive abilities of humans. On the other hands, there are many practical achievements which concentrate on game-playing agents capable of interacting with humans. In [7], Marquis *et al* have shown poker playing agents that embody emotions and communicate with people using multiple channels of communication. In [8], Kovács *et al* have built a virtual character to play chess on a physical board, which can express different emotions and produce speech.

In spite of many challenges from various angles mentioned above, because human social skills are extremely sophisticated, it is required to elaborate robots which can understand complex social skills and interact with people socio-emotionally in actual environments. So, in this study, we considered that understating deception in terms of mind reading plays a key role in complex and realistic interactions for social robots. Normally, when detecting deception through nonverbal contexts, the accuracy rate is usually about 50 percentages. However, raising stakes has an influence on the motivation to succeed in deceiving, and detecting the nonverbal deception in high stakes is easier than in low stake[9], [10]. In order to present how a humanoid robot can understand human deceptions in real social situations, we chose high-stakes poker since deception is inherent in game-playing. The humanoid plays poker, manipulating actual cards and observing human behaviors to understand how human feels and what human intents in the game. For mind-reading, our robot interprets nonverbal deception cues from multimodal stimuli such as facial expressions, eye blink, gestures and vocal stress. By observing human behaviors, it can realize what human reveals in the interaction.

In this paper, Section 2 will introduce the system overview for the social interaction. Section 3 will describe nonverbal deception cues that occur in daily social interaction and visual and auditory perceptions that have been implemented on the humanoid robot. Subsequently, the game environment will be briefly explained in Section 4 and the experiment will be demonstrated in Section 5.

2 System Overview

We formulated a component structure for the human-robot social interaction as shown in Figure 1. The humanoid perceives situational stimuli such as cards and their location. Also it observes emotional states of human player through visual and auditory sensors. The recognized information about current situation is memorized as experiences and considered when the humanoid selects next actions by decision-making based on present given situation and past experiences. The selected actions can have an effect on the surroundings including human and game, especially on the human player's circumstantial judgment and decision-making. This paper focuses on the groundwork for developing the ability to perceive and the ability to manipulate the surroundings.

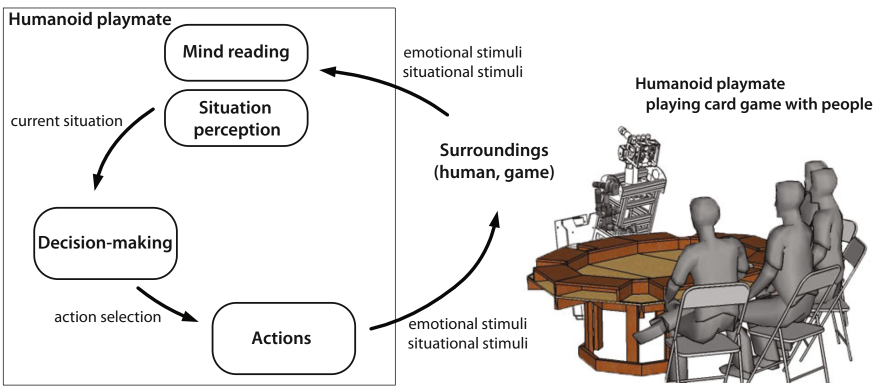


Fig. 1. A schematic diagram of human-robot social interaction

3 Nonverbal Cues for Mind Reading

3.1 Facial Feature Detection for Facial Expression Recognition

Much research on the body in cognitive sciences insists that temporal dynamics of human facial behavior such as the timing and duration of facial actions are a critical factor for interpretation of the observed behavior[11]. In particular, automatic facial expression recognition related research emphasizes the importance of facial expression's temporal dynamics for deception detection[12]. In this research, we adopted Active Appearance Models (AAMs) which are widely used for tracking and analyzing facial expressions.

AAM is a method to detect an object in 2D images using the statistical representation of shape and appearance variation[13]. AAMs align the pre-defined linear shape model with linear appearance variation to a previous unseen image containing the object. AAMs generally fit their shape and appearance components through a gradient decent. The appearance variation is linearly modeled by Principal Component Analysis (PCA) of shape $\mathbf{s}$ and texture $\mathbf{g}$.

$$\mathbf{s} = \mathbf{s}_m + \phi_s \mathbf{b}_s \qquad \mathbf{g} = \mathbf{g}_m + \phi_g \mathbf{b}_g \tag{1}$$

where $\mathbf{s}_m$, $\mathbf{g}_m$ are the mean shape and texture respectively, ϕ_s, ϕ_g are the eigenvectors of shape and texture covariance matrices. A third PCA is performed on a concatenated shape and texture parameters $\mathbf{b}$ to obtain a combined model vector $\mathbf{c}$.

$$\mathbf{b} = \phi_c \mathbf{c} \tag{2}$$

From the combined appearance model vector $\mathbf{c}$, a new instance of shape and texture can be generated.

$$\mathbf{s}_{model}(c) = \mathbf{s}_m + \mathbf{Q}_s \mathbf{c} \qquad \mathbf{g}_{model}(c) = \mathbf{g}_m + \mathbf{Q}_g \mathbf{c} \tag{3}$$

AAM fitting consists of minimizing the difference between the closest model instance and the target image by solving a nonlinear optimization problem. We employed Inverse compositional fitting methods.

3.2 Head and Hand Tracking for Gesture Recognition

According to Interpersonal Deception Theory, the use of gesture can lead to misinterpretations. Some researches have reported that when people lie, they display fewer of the gestures[15]. We considered that deceptive gestures naturally come out during the poker gaming, so that head and hand movement is one of influential factors in deception detection in the game.

For implementing head and hand movement tracking, it was assumed that in the captured images, humans move actively rather than backgrounds. Also, it was supposed that the humanoid's head movement to explore the environment will cause dynamic background changes. In order to extract foreground which indicates human's movements, we used a mixture of Gaussians to perform background subtraction in color images. A mixture of K Gaussian distributions adaptively models each pixel color. The probability density function of the kth Gaussian at pixel (i, j) at time t can be expressed as

$$N(x_{i,j}^t | m_{i,j}^{t,k}, \Sigma_{i,j}^{t,k}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \tag{4}$$

Fig. 2. Facial expression tracking using Active Appearance Models

where $x^t_{i,j}$ is the color of pixel (i, j), $m^{t,k}_{i,j}$ and $\Sigma^{t,k}_{i,j}$ are the mean vector and the covariance matrix of the kth Gaussian in the mixture at time t respectively. Each Gaussian has an associated weight $w^{t,k}_{i,j}$ (where $0 < w^{t,k}_{i,j} < 1$) in the mixture. The covariance matrix is assumed to be diagonal to reduce the computational burden, that is, $\Sigma^{t,k}_{i,j} = diag((\sigma^{t,k,R}_{i,j})^2, (\sigma^{t,k,G}_{i,j})^2, (\sigma^{t,k,B}_{i,j})^2$ where R, G and B represent the three color components.

A K-means approximation of the EM algorithm is used to update the mixture model. Each new pixel color value, $x^t_{i,j}$, is checked against the existing K Gaussian distributions, until the pixel matches a distribution. The rest of our implementation followed the background subtraction technique in [16]. After the background subtraction, median filtering is performed to remove noise.

Next, the detected foreground pixels are processed further by skin color segmentation. Skin color segmentation is performed using the YCrCb color space. Color is represented by luma computed from nonlinear RGB, constructed as a weighted sum of the RGB values, and two color difference values *Cr* and *Cb* that are formed by subtracting luma from RGB red and blue components[17].

The output image extracted by the background subtraction and skin color detection includes potential candidates for the coordinates of head and hand. The face location is separated by haar-like feature based tracking to facilitate the segmentation between head and hand. Consequently, the hand in the image which the face is excluded can be easily detected by tracking blob features.

3.3 Eye Blink Detection

Some researches have shown that the eye blinking rate decreases when the cognitive load is increased. The measure of blink rate could provide another clue for the detection of deception[18].

At first, we took a template matching based approach to find the eye location. The basic idea of template matching is that one has the template of the sought feature region that has a high similarity with other images of that feature. The template and the similarity measurement are major parts within template matching. The right eye template is used in our system considering that the blinking of both eyes occurs simultaneously. A judgment of eye is made according to the similarity between the input image and the template. After seeking the most similar region with the right eye template, the eye blink is be detected.

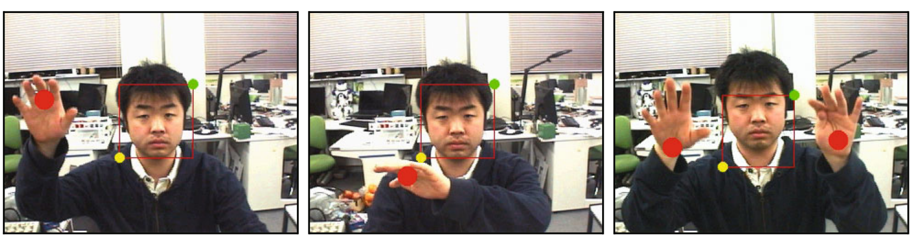


Fig. 3. Head and hand tracking for gesture recognition

Fig. 4. Blink detection

Each input image containing the closed eye and the opened eye has different similarity value as compared with the template image.

3.4 Voice

When we feel being deceived by people, we pay attention not only to their bodily expressions but also to vocal information such as voice tone and voice pitch. According to research on stress and lie detection[19], non-invasive physiological features like voice pitch variation is correlated to high stress situations. We have realized the basement to analyze the deceptive vocal information by MATLAB. In this paper, the spoken voice during the poker playing was recorded.

4 Humanoid Playmate

When a game is too easy, people feel bored and when it is too difficult, they feel frustrated. Balancing game with different levels is one of key issues in game design. From the psychological perspective of manipulating mind, one of our game strategies is to keep people to play the game for satisfaction and happiness. Game level adjustment plays an important role in satisfying this requirement. In poker game strategies, we would follow the flow model (see Figure 5), defined by M. Csikszentmihalyi as the mental state in which people are involved in an activity that nothing else seems to matter. Another game strategy is that the humanoid learns and imitates the human player's behaviors to create a sense of interacting with life-like machine and drive the human's affective responses in determining the behavioral reaction.

In order to accomplish the strategies, we built the humanoid as a playmate shown in Figure 6, named Genie developed by Artificial Intelligence Lab. at University of Tsukuba. The humanoid is composed of the upper torso with a wall-mounted 3DOF waist. It has totally 27DOF (8DOF for the head, 3DOF for the waist, 7DOF for the right arm, 4DOF for the right hand, and 5DOF for the left arm) in the body. The SSSA-Tsukuba artificial hand is comprised of three fingers and a tendon-driven mechanism. Additionally, the robot hand has a 1 DOF thumb for the adduction and abduction actuation of the thumb.

During the poker playing, the humanoid can manipulate actual poker cards. We attached an electromagnet on the fingertip of the robot hand because the

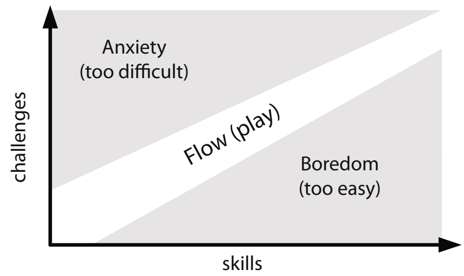


Fig. 5. Flow model with regard to challenge level and skill level

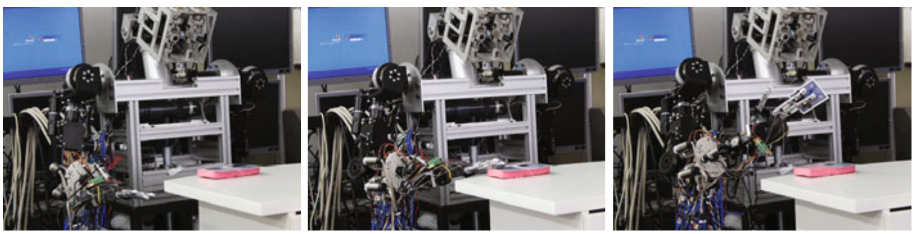


Fig. 6. Card manipulation (initial pose, reaching and picking up)

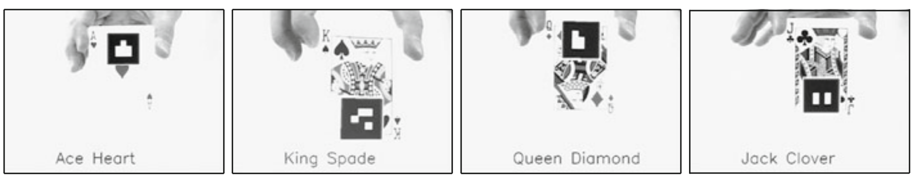


Fig. 7. Card suit recognition by AR toolkit marker

complex dynamic analysis of object manipulation by multi-fingered robot hand is out of our research scope. For physically manipulating cards, the humanoid has the kinematic profile of the movement. With the profiles, it can reach the target cards on the table as well as pick up to check its own card suits. Figure 6 illustrates the card manipulation by the right arm and hand which are initial pose, reaching and picking up, respectively. Perceiving card suits is achieved by tracking AR toolkit markers pasted on each card as shown in Figure 7.

5 Experiment

The purpose of this experiment is to evaluate the humanoid's observations on the natural scenes in a poker game regardless of deception detection and to get ideas for designing the robotic actions to measure. For the experiment, we asked two

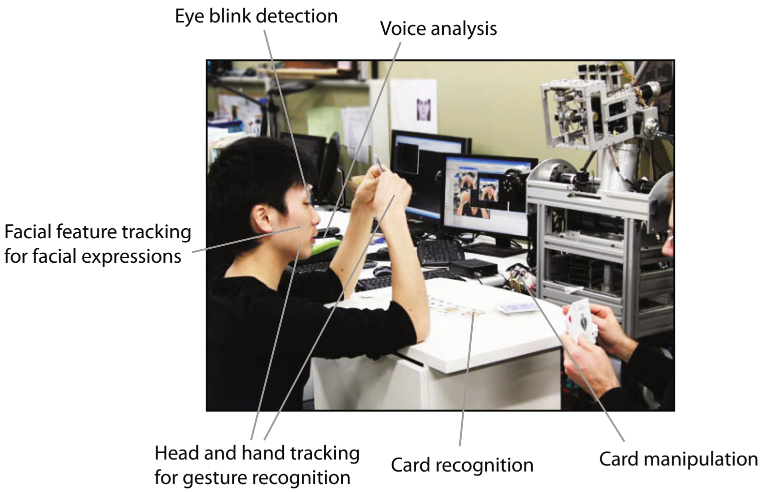


Fig. 8. Bodily expressions and voice observation during the poker game

people to play the Texas hold'em in front of the humanoid and to act naturally as shown in Figure 8. In order to encourage the participants to play for win, we provided a incentive. We then made the robot observe the behaviors of the subject who sat before it when two players were playing the game. It perceived his facial expressions, gestures, eye blink and voice without moving its body. The facial features and voice were recorded during the entire poker game. Because we expected that the player would not perform gestures and eye blinking a lot according to [18] and [20], they were analyzed by turns. We had implemented the card suit recognition and the card manipulation as described in Section 4. However, they were not applied to this experiment since the Texas hold'em game engine is not incorporated into the humanoid.

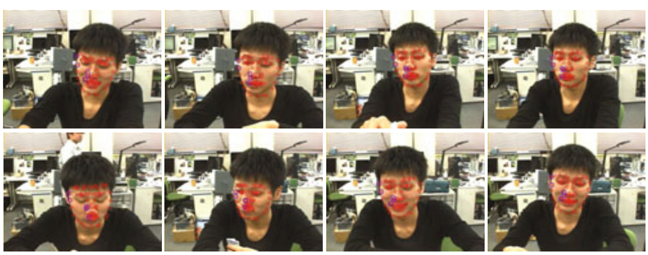


Fig. 9. Detection of facial features

Figure 9 demonstrates the facial features detected by using Active Appearance models. The 68 red landmarks fit the change of detected facial features. During the game, although the subject moved and turned the head a lot, the facial features were easily detected within the possible measurement range.

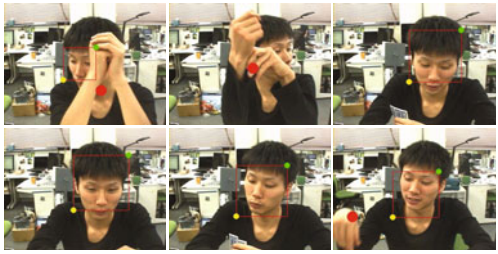


Fig. 10. Head and hand tracking

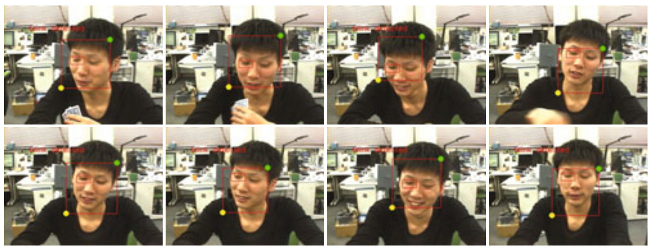


Fig. 11. Eye blink detection

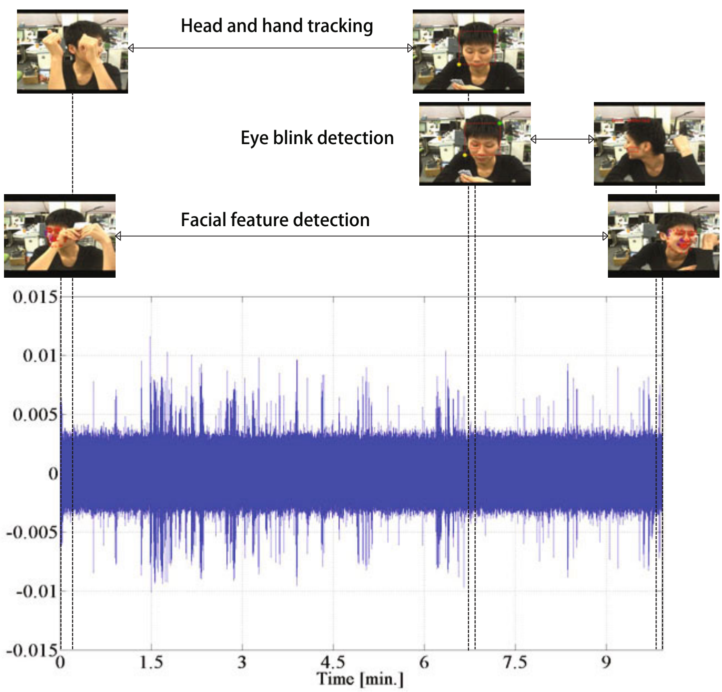


Fig. 12. The recorded sound wave in the time domain

In the game, the subject did not move his arm and hand much except for picking up of fake money for betting and shuffling the cards as a dealer. On the other hand, he often moved his head to comprehend the situation, and moved when he laughed out. The hand movements were traced excluding the cases that the subject put the hands on the table out of the range of vision. As shown in Figure 11, for the natural behaviors, the eye blink was not detected well because it was not easy to distinguish between the real eye blinking case and the other cases where the eyes were seemed to be closed when the subject looked down without hanging the head or really closed when he smiled broadly.

The sound which was occurred during the whole game is shown in Figure 12. The figure demonstrates the sound wave plotted in the time domain. The input raw sound was preprocessed by a band-pass filter within the voice frequency to reduce noise in the environment.

6 Conclusion

In this research, we have designed a humanoid playmate to understand human mind in terms of mind reading. Especially, a perceptual system has been implemented to acquire the circumstantial information such as nonverbal expressions and the poker cards. Also, simple card manipulation has been achieved simply to reach, pick up and put down the cards. In order to analyze the deceptive vocal information, we have realized a basic sound processing module. Finally in the experiment, the human player's natural behaviors were observed during a real poker game.

In the near future, we would improve the humanoid's actions to observe actively by moving the head, waist etc because the static observation has a limitation of analyzing the dynamic human behaviors. Additionally, in order to play with people interactively, we would integrate the Texas hold'em engine into the humanoid and develop the mind reading mechanism by combining the bodily expressions and voice to understand intentions and feelings.

References

1. Wimmer, H., Perner, J.: Beliefs about beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception. Cognition 13, 103–128 (1983)
2. Baron-Cohen, S., Tager-Flusberg, H., Cohen, D.: Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience, Oxford (1999)
3. Homaz, A.L., Berlin, M., Breazeal, C.: Robot Science Meets Social Science: An Embodied Computational Model of Social Referencing. In: COGSCI 2005 Workshop, pp. 7–17 (2005)
4. Thomaz, A.L., Breazeal, C.: Robot Learning via Socially Guided Exploration. In: 6th IEEE International Conference on Developmental Learning, pp. 82–87 (2007)
5. Dautenhahn, K., Nehaniv, C.L., Walters, M.L., Robins, B., Kose-Bagci, H., Assif Mirza, N., Blow, M.: KASPAR - A Minimally Expressive Humanoid Robot for Human-Robot Interaction Research. Special Issue on Humanoid Robots, Applied Bionics and Biomechanics, 369–397 (2009)

6. Kozima, H., Michalowski, M.P., Nakagawa, C.: Keepon: a Playful Robot for Research, Therapy, and Entertainment. Int. J. of Social Robotics 1, 3–18 (2008)
7. Marquis, S., Elliott, C.: Emotionally Responsive Poker Playing Agents. In: Notes for the 12th National Conference on Artificial Intelligence Workshop on Artificial Intelligence, Artificial Life, and Entertainment, pp. 11–15 (1994)
8. Kovács, G., Ruttkay, Z., Fazekas, A.: Virtual Chess Player with Emotions. In: 4th Hungarian Conference on Computer Graphics and Geometry (2007)
9. DePaulo, B.M., Stone, J.I., Lassiter, G.D.: Deceiving and Detecting Deceit. The Self and Social Life, 323–370 (1985)
10. DePaulo, B.M., LeMay, C.S., Epstein, J.A.: Effects of Importance of Success and Expectations for Success on Effectiveness at Deceiving. Personality and Social Psychology Bulletin, 14–24 (1991)
11. Ambadar, Z., Schooler, J.W., Cohn, J.F.: Deciphering the Enigmatic Face: the Importance of Facial Dynamics in Interpreting Subtle Facial Expressions. Psychol. Sci. 403–410 (2005)
12. Valstar, M.F., Pantic, M., Ambadar, Z., Cohn, J.F.: Spontaneous vs. Posed Facial Behavior: Automatic Analysis of Brow Actions. In: 8th International Conference on Multimodal Interfaces, pp. 162–170 (2006)
13. Matthews, I., Baker, S.: Active Appearance Models Revisited. Int. J. of Computer Vision 60, 135–164 (2004)
14. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 484–498. Springer, Heidelberg (1998)
15. Vrij, A., Edward, K., Roberts, K.P., Bull, R.: Detecting Deceit via Analysis of Verbal and Nonverbal Behavior. J. of Nonverbal Behavior, 239–263 (2004)
16. Javed, O., Shafique, K., Shah, M.: A Hierarchical Approach to Robust Background Subtraction using Color and Gradient Information. In: Workshop on Motion and Video Computing, pp. 22–27 (2002)
17. Vassili, V.V., Sazonov, V., Andreeva, A.: A Survey on Pixel-Based Skin Color Detection Techniques. In: Proc. Graphicon-2003, pp. 85–92 (2003)
18. Fukuda, K.: Eye Blinks: New Indices for the Detection of Deception. Int. J. of Psychophysiology, 239–245 (2001)
19. Sung, M., Pentland, A.: PokerMetrics: Stress and Line Detection through Non-Invasive Physiological Sensing. Ph.D. thesis, MIT Media Laboratory (2005)
20. Caso, L., Maricchiolo, F., Bonaiuto, M., Vrij, A., Mann, S.: The Impact of Deception and Suspicion on Different Hand Movements. J. of Nonverbal Behavior 30, 1–19 (2006)

DreamThrower: Creating, Throwing and Catching Dreams for Collaborative Dream Sharing

Noreen Kamal, Ling Tsou, Abir Al Hajri, and Sidney Fels

Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, British Columbia, Canada
noreen.k@ubc.ca, tsoul@ece.ubc.ca, alhajria@interchange.ubc.ca, ssfels@ece.ubc.ca

Abstract. The *DreamThrower* is a novel technology that explores virtually creating, throwing and catching dreams. It detects users' dream state by measuring rapid eye movement (REM). Once the dream state is detected, sound and light stimuli is played to alter the dream. Users report on their dream, and they can send the stimuli that they have used to another person via an on-line website. A working prototype accurately detects REM sleep. Based on preliminary results, the sound and light stimuli were found to have little influence on their dreams. Our prototype's ability to detect REM effectively coupled to a social network to share dream stimuli opens up a fun game environment even if the stimuli itself does not have a significant impact. Instead, user engagement with the social network may be sufficient to alter dreams. Further studies are needed to determine whether stimulus during REM can be created to alter dreams significantly.

Keywords: Dream detection, dream creation, dream stimulus, Rapid Eye Movement.

1 Introduction

The term *Dreamcatcher* has become a common place keepsake in North American culture. This idealized and over commercialized object originated from the North American Aboriginal Ojibwa Nation that, traditionally, hung above the bed as a charm to protect children from nightmares, as the dreamcatcher acted as a snare for bad dreams, and only let good dreams through [12]. By taking this analogy of a *dreamcatcher* and adding the concept of dream throwing and dream creating, we can start to conceive the *DreamThrower*. Have you ever wondered if you can share your dreams with others? Have you ever considered if by sharing external stimuli, others would have the same or similar dream as you? This research looks at how the knowledge of dream creation and control can be used to create a novel solution to "throw" dreams to other people.

DreamThrower is a fun game concept that alters users' dreams by detecting the dream state and playing external sound and light stimuli, and then coupled with a social network to share these experiences with their friends. Our digital *DreamThrower,* shown in figure 1, detects when a person has entered a dream state through eye

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 20–31, 2010.

movements using an infrared (IR) sensor. It uses external stimuli such as light and sound to enhance events in one's dreams. When the user is awake, they would diary the events of the dream, as well as rank how pleasant or unpleasant (valence) the dream was. A user would be able to send another person a suggested stimulus via a social network service, or *throw* their dream to another user.

This rich dream experience combines socialization of dream content and collaboration of sharing dream experiences. This is done by not only using stimuli to alter dream experiences, but also to share these experiences and stimuli with others to better understand if others will have the same or similar dreams. Our prototype's ability to detect rapaid eye movement (REM), which is a phenomenon of extremely active movement of the eye that can be observed while dreaming, effectively coupled to a social network to share dream stimuli opens up the possibility of a fun game environment even if the stimuli itself does not have a significant impact. Instead, user engagement with the social network may be sufficient to alter dreams to discuss with their network of friends.

In the next section, we discuss the related work in the area of dreams and REM, dream stimuli and technologies related to dreams. In section 3, we describe our system design. In section 4, the user evaluation of our prototype is outlined, which is followed by our results (section 6), and finally the future works and conclusion is presented.

Fig. 1. Photograph of *DreamThrower* prototype

2 Related Work

There has been significant work done in the area of dreaming from the perspective of behavioural psychology and brain science; however, the research done around the design of novel technologies in the domain of dreams is limited. Existing work includes understanding different dream states, lucid dreaming, effects of external stimuli while dreaming, and measuring physiological outputs when dreaming.

There has been research done in understanding different states when dreaming. It is widely accepted that REM strongly correlates to the dreaming state, and that REM dreaming state is qualitatively different than non-REM sleep and waking states [15].

The area of lucid dreaming is a domain that has spurred both academic and non-academic studies. The concept of lucid dreaming is structured around being aware while one is dreaming, being able to control a dream and being able to remember dreams [16, 18, 20]. Technology has been created to assist a person to be lucid when dreaming [11, 18, 20, 23]. These technologies include technologies that first detect when a person has fallen asleep such as the NovaDreamer® device [11, 20]. The DreamSpeaker® provides audible signal and the DreamLight® provides light signals to assist the sleeper to understand that they have fallen into dream state to assist them in maintaining a lucid state while dreaming [11]. Similarly, the Lucid Dream Machine assists people in understanding that they are dreaming to help them in achieving lucid dreams [18]. The REM-Dreamer is another commercial product that works on the same premise but with the added functionality of being able to adjust the intensity of dream stimuli so that it is properly correlated for an individual in a dream state [23]. These technologies use an IR sensor to detect eye movement (REM) to signal when a person is in a dream state; another study looked at methods to detect REM by non-visual means [1]. Our project will use the methods that have been commercialized and studied to detect the dream state by measuring REM and also to understand the concepts of lucid dreaming; although our project will not focus on achieving and maintaining a lucid state when dreaming, it will need to be understood, so that our participants can remember and diary their dream experiences.

Another related research area is in understanding how external stimuli affect dreams. Sound stimuli while sleeping has been studied extensively: verbal sentences during REM sleep can assist in accessing declarative knowledge during sleep and help consolidate knowledge but these verbal cue will not be inserted into dream content [10]; the insertion of verbal content when sleeping also showed that external verbal stimuli are perceived as belonging to the events of the dream [3]; the insertion of verbal stimuli prior to falling asleep was also studied and shown to effect dream content as well as recall after waking [9]; and finally another study revealed that sound stimuli while sleeping cannot be used as an alert [5]. Smell stimuli has also been studied although it has been shown that olfactory senses are very limited while sleeping, and therefore has little effect on dream content [2, 8]. Finally, the use of light has been used in the lucid dreaming domain to alert a person of a dream state [21]; therefore, we can also appreciate that light or visual stimuli have a cognitive effect while dreaming. Our project will use light and sound including verbal sound as stimuli to change the events of the dream and facilitate dream throwing.

There have also been several studies that look at measuring physiological outputs when dreaming. Various physiological signals correlate to REM or dreaming activity levels [15, 16, 22]. Hobson et al. in 2000 developed the activation-input source-neuromodulation model (AIM), where electroencephalogram (EEG) activation and firing level of reticular, thalamic and cortical neuron correlates to the activation level; level of presynaptic and postsynaptic inhibition and excitability of sensory motor patterns correlates to internal or external information source; and activity of level of aminergic neurons correlates to mode or organization of data. Holzinger et al. in 2006 studied the electrophysiological differences between lucid and nonlucid dreams in REM sleep by comparing the frequency of EEG signals. EEG signal frequency while sleeping and waking states has been stratified into 7 levels; for example, the delta level is with frequency of 0.1 to 3 Hz, which corresponds to deep sleep and lucid dreaming and the 3-8 Hz is deep relaxation compared to 40 Hz is high-level information processing [4]. Further studies in understanding EEG signals have studied to further understand the alpha level of EEG activity in REM sleep [6]. Our work will not look at physiological outputs, but rather rely on self-reports of the dream after the person wakes up.

Another related work that deserves mention is the Dream Communicator, which was one of nine conceptual design proposals for information appliances in [13]. Although the Dream Communicator was only perceived conceptually in this paper and never developed, its concept is to allow distant lovers to enter each other's dream state by stimulating their dreams with sounds or speech; this idea builds on the ability for external stimuli to enter the dream state. Although this design idea presents a portion of the *DreamThrower* of sending some stimuli to a remote person, it does not provide outputs from the person to visualize their dream experiences, nor the idea of dream crafting and sharing these creations through diary logs.

3 System Overview

The *DreamThrower* is a system that will allow a person to alter their dreams by providing selected light and sound stimuli. The stimuli will only be triggered when REM sleep is detected indicating that the user is dreaming. This unique system will also have an on-line socialization of dreams component, where users can share their dream experiences by self-reporting their dreams, and then "Throw" (or share) these stimuli with their on-line *DreamThrower* friends.

The system will also include a *DreamThrower* eye mask component. The eye mask will contain an IR emitter and detector pair to detect REM sleep. Once REM sleep is detected, the selected light and sound stimuli will start until REM sleep ends. Immediately after the dream state ends, the *DreamThrower* prototype will play a loud sound that says, "wake up", in order to wake the person who is asleep, so that they can report on their dream. Dreams will be self-reported on the *DreamThrower* website for sharing on-line with friends.

This study does not focus on lucid dreaming or try to invoke lucidity while dreaming. There has been significant work already done in lucid dreaming, so this study will not repeat this body of existing work. The focus of this work will be to invoke dreams by selecting stimuli, and to share dream experiences with others by sharing the same stimuli.

3.1 Design

The detailed design for the *DreamThrower* system is shown in figure 2. The main components of the design are the following:

1. Eye mask: contains the IR emitter and detector, the light stimulus and the sound stimulus
2. Arduino board: provides the power and the processing and control for the dream detection and playing of the stimuli, which is coded in C
3. Computer: loads the stimuli on to the Arduino board and provides the link to the on-line website.

3.2 REM Detection

To apply the stimuli at the appropriate time, the *DreamThrower* system will need to detect when a person is dreaming by detecting the REM state. In order to detect a REM state, an IR light signal will be shined across one eye towards the IR detector.

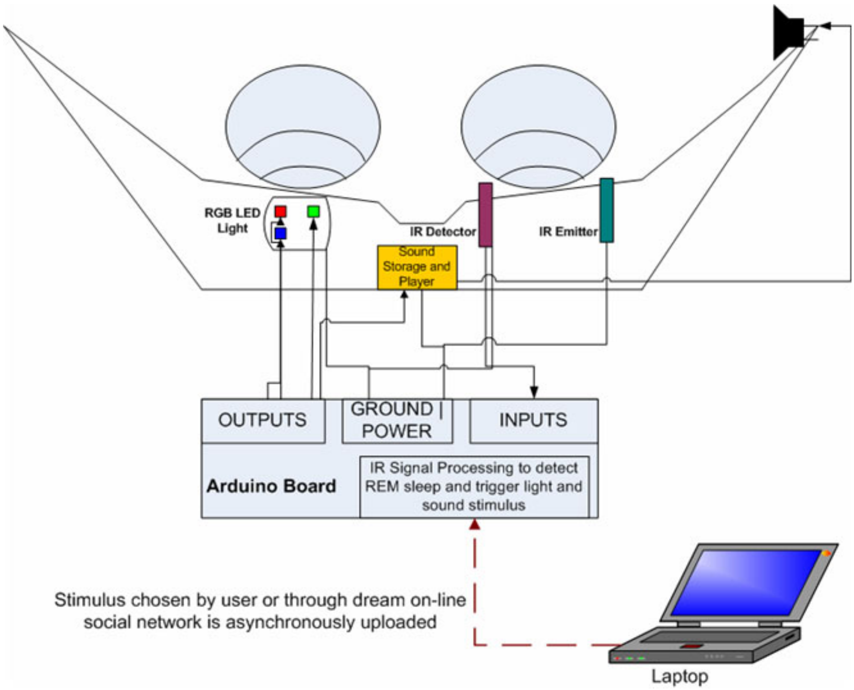


Fig. 2. Detailed design for *DreamThrower*

When the pupil of the eye is stationary, the IR light will be blocked, and when the pupil moves the IR light will be exposed towards the detector. This is shown in figure 3. The IR detector supplies different levels of voltage to the A/D converter on the Arduino board depends on the amount of IR light received. When the IR light is partially blocked, a low IR signal is detected, and when the IR light is exposed, a high IR signal is detected.

To accurately detect the REM stage, the system performs a series of signal processing algorithms on the sampled IR measurements. A preprocessing step is first done for noise reduction by a combination of windowing and integration. The basic rectangular window is used to divide the signal into blocks of data each contains the IR

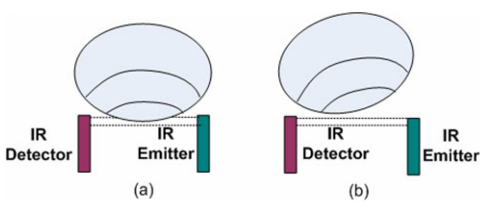


Fig. 3. REM detection using IR detector and emitter. (a) shows the partial blocking of light when eye is in the position when the eye is not moving, (b) shows the full light signal being received when the eye moves to the side.

measurements over a small time interval. These blocks are then integrated to provide a single data value for each windowed period. Undesired high frequency noise is filtered, and therefore, makes the signal generated in the REM stage more distinct compared to the base signal where the eyes are quiescent.

The entire REM detection algorithm can be divided into three stages: calibration, REM start detection and REM terminate detection. The calibration stage is used to obtain a baseline for the signal and is triggered after a preset waiting time. It is of import to not start the calibration before the user falls asleep since the eye activities of a person during the REM stage is relatively similar to the activities when he/she is resting but awake [14]. A threshold value, which is required in the later two stages, is set according to this baseline value.

Once the threshold obtained, the system is now ready to perform REM detection. The data points within a window of few seconds are used to observe the level of eye movement activities. After comparing each point with the threshold, a percentage of the number of points excess the threshold level can be calculated. A low percentage indicates that the eye was mostly stationary and pointing straight ahead whereas a high percentage indicates that the eye spent most of the time pointing away from the centre which is caused by extraocular muscle twitches during REM state.

Figure 4 shows the calculated integrated IR signals over one sleep cycle. This plot shows the pre-set waiting time, the calibration step, the non-REM sleep is shown during the low values, and the REM sleep can be seen in the figure as the high values.

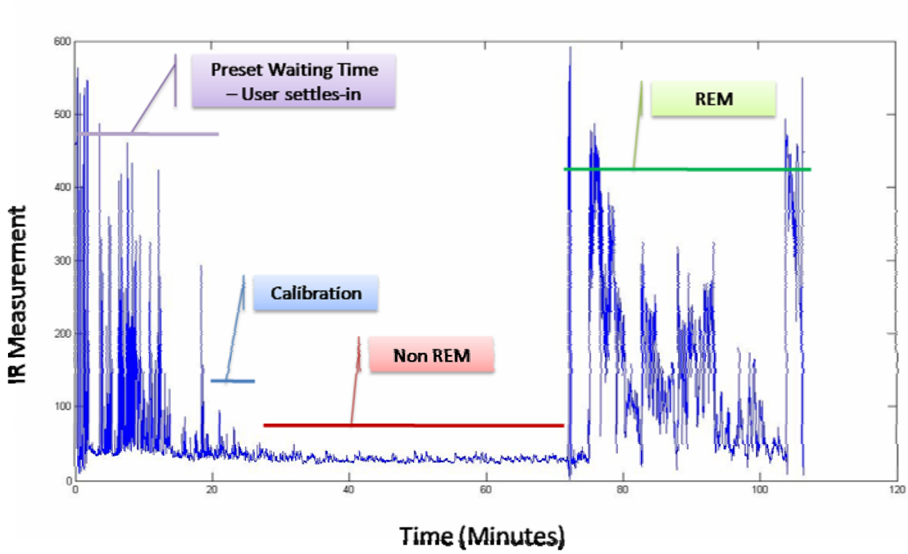


Fig. 4. Plot of integrated IR signal using the *DreamThrower* prototype. After settling in time (0-20min) we calibrate the background signal (20-25min). The subject is sleeping but not dreaming (27-70min) and then enters REM sleep (70-110min).

3.3 Sound and Light Stimuli

Once REM sleep is detected by the *DreamThrower*, light and sound stimuli that are selected by the user will be played until REM sleep ends. The light stimulus will be sent from the Arduino board to a multi-coloured LED light that will be on the eye mask. This LED light will be able to emit any light colour, which will be flashing on and off at a set frequency. This will be done by controlling the output signals of the Arduino board that are sent to the red, green and blue inputs on the RGB LED light.

The sound stimulus will be triggered when a command is sent from the Arduino board to a VMusic2 module. This VMusic2 module will continuously play the selected sound file from a connected memory stick when REM sleep is detected. The VMusic2 module will also play a sound to wake the user after the REM sleep has ended.

3.4 Prototype

A prototype for the *DreamThrower* was developed. This prototype closely followed the detailed design for the eye mask as described in Section 3 of this paper. Figure 1 shows a photograph of the prototype. Laboratory safety glasses were used for the structure of the *DreamThrower* eye mask. The IR detector and IR emitter were placed on the left eye of the goggles, and the light stimulus emitter was placed on the right eye of the goggle. The Arduino board, battery, and VMusic2 were placed on-top of the goggles.

The prototype had the following sounds available for the evaluation of the *DreamThrower* system:

- Jungle sound
- Ocean wave sound
- Heavy rain sound
- Wind sound

It is recognized that these sounds are currently limited to environmental sounds. This will provide focus to our evaluation in understanding the effects of environmental sound stimulus. Further study can be done to include other sounds such as music and verbal speech, as it has been recognized that speech sounds do influence dreams [9, 10]. Once the system detects that REM sleep has ended, the prototype plays a speech recording, “wake-up” very loudly, so that the user can awaken and report their dream experiences, which was done so that the prototype can be evaluated. It is envisioned that the user would be able to decide if they wish to be awoken after their first dream or second dream or never during a night’s sleep. This is important as sleep continuity is of particular concern.

The light stimulus could also be changed in the prototype. The RBG LED light could take any combination of signals from the Arduino program to create a multitude of colours. The current Arduino program will create a flashing light at a set frequency of the colour that is chosen. The sound and light stimuli can be changed directly in the Arduino program and uploaded to the board.

The on-line system was not developed for the prototyping of the system; however, the self-report questions were asked using a questionnaire in the evaluation experiments.

4 User Evaluation

This prototype was used for the user evaluation of the *DreamThrower* system. The purpose of the evaluation was to answer the following research questions:

1. Does the *DreamThrower* system accurately detect dreaming state while sleeping and play the stimuli while dreaming?
2. Does the user of *DreamThrower* feel that the system influenced his/her dreams?
3. Did the user feel that the *DreamThrower* system could be fun to use and share stimuli?

4.1 Evaluation

The system was tested using 3 subjects, who were outside of the research team. The small number of subject for this evaluation was to establish an initial foundational understanding of the *DreamThrower* system. There were one male and two female subjects.

Each subject participated in 1 experiment. Once again this was to obtain a foundational understanding of the influence of dreams using the stimuli. Table 1 show the type of stimuli that were used on each subject.

Table 1. Stimuli that were used on each subject

Subject	Gender	Sound	Light
Subject 1	Female	Jungle	Green
Subject 2	Male	Jungle	Green
Subject 3	Female	Ocean	Blue

In order to answer the first research question, data was also collected using the Arduino program to help facilitate the evaluation of the system. This included a flag to confirm if the user entered REM sleep and a flag to confirm that the sound and light stimuli was played. This was confirmed by asking the participants if they just had a dream prior to waking-up.

The subjects were informed which sound and light stimuli will be played while they are dreaming. Subjects were informed of the stimuli as this is in line with the usage scenario for *DreamThrower,* where users download the stimuli onto the device. The eye mask was calibrated for each subject to ensure that the REM detection would function properly for their eye shape. They would wear the mask while sleeping, and when they wake up, each subject was asked to complete an initial questionnaire. This questionnaire inquired if they had a dream and how much of it they could remember.

If the subject did have a dream and remember some of it, they were asked to complete a self-report questionnaire. This questionnaire contained the same inquiries as the on-line self-report page. This will provide an understanding of the dream that the subjects had during the experiment, and it will also provide an analysis of the questions that were chosen for the self-report page.

Finally, each subject was interviewed to obtain rich data set to qualitatively understand the subjects' thoughts about the *DreamThrower* system and the amount of influence that it had on their dreams. The purpose of the questionnaires and interview questions were to answer the second and third research questions.

4.2 Results

Based on the data retrieved from the Arduino program, we were able to determine that the *DreamThrower* prototype was able to accurately detect REM sleep and play the selected stimuli for each of the three subjects. This was confirmed by the plot of the IR integrated signal for each user and correlating this with the flag of when the stimuli was played.

The results of each subject's retention and influence of the stimuli on their dream is summarized in table 2.

Table 2. Results of user evaluation

Subject	Remember Dream?	Noticed Stimuli?	Felt their dream was influenced by stimuli?	Subject of dream
Subject 1	Little	Yes, sound and light	No	TV Series
Subject 2	Very little	Sound only	No	Sports highlights
Subject 3	Very little	No	No	*DreamThrower device*

Although the amount of dream content that the subjects remembered varied from little to very little, all three subjects can recall that they had a dream, and they had some recollection of the content of the dream.

The awareness of the stimuli varied for each of the subjects. Subject 1 did notice the light and sound stimuli while dreaming and was awakened by them. The subject 2 only noticed the sound but not the light stimulus, where as subject 3 did not notice any stimuli at all. In the case where the subjects did notice the stimuli, the stimuli were treated by the subjects as actual phenomenon happening in the real world rather than in their dreams. None of the subjects felt that the content of the dream was affected by the stimuli.

No correlation can be observed between the contents of the dreams of the three subjects: subject 3 dreamed about the *DreamThrowe*r device itself whereas subject 1 dreamed about a TV series that she watched shortly before starting the experiment, and subject 2 dreamed about sport highlights. All three subjects rated their dreams to be neither disturbing nor pleasant.

The interviews with the subjects also revealed some thoughts on their perception of the level of influence that the device had on their dreams. One subject felt that he/she would like to test the device multiple times: *"I think, you have to have a series of experiments in your subjects because it is not easy to get familiar with the device as I was affected by the device itself and not by the stimuli"*.

Another subject commented about the comfort of the device itself, which speaks to how the device needs to be improved: *"The device itself is too uncomfortable to wear. Also, I was awaken by the stimuli shortly after I started dreaming."*

When the subjects were asked if they found the device and the concept of altering dreams based on stimuli "enjoyable", all three answered "no". Although, two of the three subjects said that it was because the device did not alter their dreams, which suggests that if the device could alter the content of their dreams, they may have found it "enjoyable" to use.

5 Discussion

The data that was collected from this preliminary user evaluation confirms that the *DreamThrower* device does accurately detect the dream state of people while they are sleeping. This is because it accurately detected the REM state of all three of the experiment subjects. Therefore, we can confirm our first research question: the *DreamThrower* system does accurately detect dreaming state while sleeping and play the stimuli, while dreaming.

However, it is evident that our data set on how sound and light stimuli influence dreams is very limited. Therefore, no hard conclusion can be drawn on whether *DreamThrower* is able to influence dreams. In order to fully understand how environmental sound influence dreams, we would need to test each subject with each sound and light combination with varying volumes and light intensity as well as light flashing frequency. Additionally, it is evident that each subject had very different interaction with the stimuli when it was play. This suggests the high variability in people to filter out external sounds and light when dreaming. Therefore, a user evaluation on a larger number of subjects is needed to fully understand how the *DreamThrower* device influences dreams.

It is evident, however, that all three subjects' dreams were not influenced by the device. Some possible reasons why the device did not have an effect on the subjects' dreams:

1. The type of audio/light stimuli chosen was not easily acceptable by humans as the content of their dreams. For example, using a recorded conversation instead of natural sounds may have a greater impact on the subject's dream.
2. For the subjects to interpret the applied stimuli as a part of their dreams might take some practice since humans tends to react (wake up when hearing the alarm clock) or filter out (sleep through the sounds of a thunder storm) external stimulation during sleep.
3. The level of audio volume and light intensity might have an effect. If the volume or light intensity was too high, the subject was likely to be awakening by the stimuli before they have any effect on the dream. However, if the volume or light intensity was too low, the subject would not even have noticed the stimuli.

Furthermore, it is possible that the stimuli did have an effect on the lucidity of the subjects' dreams since all three subjects were able to recall that they in fact had a dream. However, further experiments need to be performed to support this argument.

It is of importance that one of the subjects had a dream about the *DreamThrower* device, as it shows that dreams can be influenced by external stimulus. This results shows validity to the idea that dreams can be influenced by stimuli and the concept of *DreamThrower* shows promise. Additionally, it is also interesting to observe that the content of each subject's dream was based on the individual's activities prior to sleeping and also on their personal interests. Therefore, the stimuli should perhaps be more personalized to the individual rather than being limited in scope.

6 Future Work and Conclusion

Based on the results of the preliminary user study, it is evident that the idea of dream sharing and dream influence presents an interesting and entertaining concept. It is clear that more work needs to be done to better understand if sound and light stimuli can influence dream. The use of only environmental sounds, such as ocean and rain, is easy for user to filter out or to understand as external sounds. Therefore, further testing on different sounds such as music and speech sounds may yield different results. This presents an area for future evaluation of the device.

Testing the system on more users, where each user goes through the test several time with different stimuli is needed to better understand which sound and light human subjects would find to influence their dreams the most. Allowing users to develop their own stimuli based on their own interests would also provide an interesting study, as the content of the dreams are personalized to people's individual context, thoughts and activities.

The study focused on the *DreamThrower* device and only touches on the self-reporting component of the whole system. Therefore, future work should also conceptualize and design the on-line system to diary dreams, collaborate with friend and create/choose sound and light stimuli. Additional studies to understand the placebo effect of providing pre-sleep stimuli through a rich social network.

The extension of entertainment to dreaming by incorporating a rich social network environment to share dream experiences and pre-sleep stimulus presents a novel concept. Additionally, The concept of altering dreams by detecting dream state and providing stimuli is novel and presents a challenging yet interesting concept. By understanding the types of stimuli that will alter the dreams of the majority of people, a new human interface paradigm can emerge. For these reasons, it is worth further study.

References

1. Agarwal, R., Takeuchi, T., Laroche, S., Gotman, J.: Detection of Rapid-Eye Movements in Sleep Studies. IEEE Transactions on Biomedical Engineering. 52(8), 1390–1396 (2005), doi:10.1109/TBME.2005.851512
2. Badia, P., Wesensten, N., Lammers, W., Culpepper, J., Harsh, J.: Responsiveness to Olfactory Stimuli Presented in Sleep. Physiology & Behavior 48, 87–90 (1990)
3. Berger, R.J.: Experimental Modification of Dream Content by Meaningful Verbal Stimuli. British Journal of Psychiatry 109, 722–740 (1963)
4. Brain-Wave Machine (1999), `http://www.hackcanada.com/homegrown/wetware/brainwave` (Retrieved on January 31, 2010)

5. Burton, S.A., Harsh, J.A., Badia, P.: Cognitive Activity in Sleep and Responsiveness to External Stimuli. Sleep 11(1), 61–68 (1988)
6. Cantero, J.L., Atienza, M., Salas, R.M.: Spectral Features of EEG Alpha Activity in Human REM Sleep: Two Variants with Different Functional Roles? Sleep 23(6), 1–5 (2000)
7. Carskadon, M.A., Dement, C.W.: Normal Human Sleep: An Overview. In: Kryger, M. (ed.) Principles and practice of sleep medicine, pp. 16–25. W. B. Saunders Co., Philidelphis (1994)
8. Carskadon, M.A., Herz, R.S.: Minimal Olfactory Perception During Sleep: Why Odor Alarms Will Not Work for Humans. Sleep 27(3), 402–405 (2004)
9. Cipolli, C., Fagioli, I., Mazzetti, M., Tuozzi, G.: Incorporation of presleep stimuli into dream content: evidence for consolidation effect on declarative knowledge during REM sleep? Journal of Sleep Research 13, 317–326 (2004)
10. Cipolli, C., Bolzani, R., Tuozzi, G., Fagioli, I.: Active processing of declarative knowledge during REM-sleep dreaming. Journal of Sleep Research 10, 277–284 (2001)
11. Courtney, E.: CEREBREX: DreamWare Specification Document (1997), `http://www.cerebrex.com/dream.htm#dreamspk` (Retrieved on January 29, 2010)
12. Dreamcatcher (2010), `http://en.wikipedia.org/wiki/Dreamcatcher` (Retrieved on January 29, 2010)
13. Gaver, B., Martin, H.: Alternatives: Exploring Information Appliances through Conceptual Design Proposals. In: CHI 2000, April 1 - 6, pp. 209–216 (2000)
14. Herman, J.H., Barker, D.R., Roffwarg, H.P.: Similarity of Eye Movement Characteristics in REM Sleep and the Awake State. Psychophysiology 20(5), 537–543 (1983)
15. Hobson, J.A., Pace-Schott, E.F., Stickgold, R.: Dreaming and the brain: Toward a cognitive neuroscience of conscious states. Behavioral and Brain Science 23(6), 793–842 (2000)
16. Holzinger, B., LaBerge, S., Levitan, L.: Psychophysiological Correlates of Lucid Dreaming. Dreaming 16(2), 88–95 (2006)
17. Kahan, T.L., LaBerge, S.: Lucid dreaming as metacognition: Implications for cognitive science. Consciousness and Cognition: An International Journal 3(2), 246–264 (1994)
18. Ketyung: Lucid dream machine lets you have full control of your dreams (2008), `http://www.techchee.com/2008/06/20/lucid-dream-machine-lets-you-have-full-control-of-your-dream/` (Retrieved on January 29, 2010)
19. LaBerge, S.: Lucid dreaming: Evidence and methodology. Behavioral and Brain Science 23(6), 962–964 (2000)
20. LaBerge, S., Rich, R.S., Wright, D.K., Kottke, D.G.: U.S. Patent Number 5,507,716. Washington, D.C. US Patent and Trade Office (1996)
21. The Lucidity Institute: Lucid Dreaming (2007), `http://www.lucidity.com` (Retrieved on January 31, 2010)
22. Perry, E.K., Piggott, M.A.: Neurotransmitter mechanisms of dreaming: Implication of modulatory system based on dream intensity. Behavior and Brain Science 23(6), 990–1092 (2000)
23. REM-Dreamer - Lucid Dream Induction Device, `http://remdreamer.com/index.php` (Retrieved on January 31, 2010)

Everyone Can Do Magic: An Interactive Game with Speech and Gesture Recognition

Chris Wang, Zhiduo Liu, and Sidney Fels

Department of Electrical and Computer Engineering
University of British Columbia
Vancouver, BC Canada V6T 1Z4
{chrisw,zhiduol,ssfels}@ece.ubc.ca

Abstract. This paper presents a novel game design that allows players to learn how to cast magic spells that combine hand gestures and speech. This game uses the imperfect recognition performance in speech and gesture recognition systems to its advantage to make the game challenging and interesting. Our game uses a Wii remote encased in a wand and a microphone to track player's gestures and speech which are then recognized to determine if they have performed the spell correctly. Visual feedback then provides confirmation of success. Through the game, players learn to adjust their speaking and movement patterns in order to meet the requirements of the recognition systems. This effectively mimics the characteristics of casting spells correctly such that players are trying to adjust their performance so that an "oracle" recognizes their speech and movement to have a magical outcome. A user study has confirmed the validity of the idea and establishes the accuracy required to create an interesting game based on the theory of channels of flow.

Keywords: Interactive game, speech recognition, gesture tracking system, magic.

1 Introduction

Computer game design has been long investigated through the decades. The main goal of game design usually focuses on engagement and entertainment. The ideal achievement of a role playing game or an experimental learning game is to make the player experience a growth in skill and a series of discoveries throughout the gameplay. The main aspects involved in accomplishing this goal are game plot, challenge level, and game interaction [1]. An attractive game plot is the first part of a successful game design. Inspired by the popularity of magic related entertainment products such as books, movies and TV shows, developing a wizard game was our initial motivation of this work. Motivated by the Harry Potter book series [2][3], we conveyed a preliminary survey of 25 participants mostly consisting of graduate students aged between 23 and 35 years old, located across the world including Canada, USA, Singapore, and Norway (14 females and 11 males). Out of all participants, 17 of them were Harry Potter fans and they all showed great interest in games relating to this popular story plot. Additionally, 6 out of the remaining 8 non-Harry Potter fans were also

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 32–42, 2010.

willing to try a magic-themed game. Among all the participants, 80% thought a spell learning activity in such games met their expectations. This large population of Harry Potter fans in the world provides great potential for this design.

A magic spell consists of a synchronized gesture made with a verbal command. Building a wizard game capturing these features of a spell requires a system with both speech recognition to identify the spell's incantation and gesture recognition to detect the player's motions. The recognition needs to be synchronized as both correct speech and gesture are simultaneously necessary for a successful spell.

Speech recognition has been an active research area for over two decades. Recent work has shown success for speaker dependent, small vocabulary speech recognition system such as voice dialing on cell phones [4], however, due to the complexity of the human language, the accuracy for general speech recognition is still not perfect. Gesture tracking also faces similar problems. Despite being in development for years, there are no non-encumbering systems that can report positions with perfect accuracy and position, while not constricted by various interfering media. To greatly oversimplify, the continuous works are encouraging, but we are still waiting for improved speech and gesture recognition technology to be created.

Yet, the imperfectness within these technologies can be incorporated into our game as challenges, which can actually benefit the enjoyment of games. As shown in the three channel model of game design [5], challenge is an essential element of an attractive game. It was observed that the unpredictable nature of magic matches well with the flaws of these two technologies. The reasonable error rate in the speech and gesture recognition systems can be used to impose challenges on the players. The players are then required to develop their skills throughout the game to overcome the technological difficulties in order to accomplish in game goals.

The remainder of the paper is organized into four sections. Section 2 presents the background about game design and section 3 explains the implemented system. Next, user study results and discussion are presented in Section 4 and future work is presented in Section 5.

2 Background

2.1 Game Design

Studies of experimental learning games by Kiili el. introduced a three channel model to describe game design [5]. Figure 1 shows the diagram referred to as a three channel model of flow. The axes of the diagram are challenges and skills, which are the most important dimensions of the flow experience. As P represents a player, the state of a player may fall into boredom region when challenges are significantly lower than the player's skill level. On the other hand, if the challenges that a player faces are beyond their capabilities, they may feel of anxious, which forms the upper region of the diagram. Flow emerges in the space between anxiety and boredom, which implies that challenges players face in the game closely match their skill level. A well designed game should keep the player within the flow state, where both the player's skill level and challenge level grow at the same pace. Additionally, the flow channel can be extended by providing more guidance to the player or by providing the possibility of solving problems. Our design is able to accomplish this control of flow experience by providing variety of spells with different difficulties.

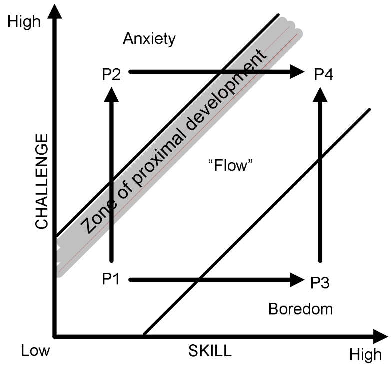


Fig. 1. Three channel model of flow

2.2 Related Work

2.2.1 Interactive Games

We reviewed the literature, and did not find a game which allows the player to be involved with both their speech and body movements except for the unreleased Kinect by Microsoft, which is advertised to have capabilities of natural speech and gesture interaction with players [6]. Although there are many video games related to magic or based on the Harry Potter books and films, such as the Harry Potter video game series developed by EA games and Lego Harry Potter games developed by Traveller's Tales, most are restricted to the visual interface only [7]. The most recently released EA game - Harry Potter and the Half Blood Prince allows very limited simple gesture interactions using the Wii remote when played on the Wii platform [8] but does not incorporate any speech interaction.

2.2.2 Gesture Recognition

The common gesture recognition technique is to equip the hand with a number of sensors which provide information about hand position, orientation, and flex of fingers. Recently, different researchers [9] have congregated to a new device for this problem: the Wii Controller. The key hardware contained in the Wii Controller that has attracted the researchers is its 3-axis accelerometer: Analog Device ADXL330. [10] This hardware emits a timed triplet $(a_x, a_y, a_z)_t$ which represents the acceleration in X,Y and Z axis at sample rates up to 80 Hz. [11]

A recent work on gesture recognition [10] has shown that a 97% recognition rate on user dependent study and 72% recognition rate on user independent study can be obtained using Dynamic Time Wrapping on ten different gestures. The gestures used in that work are similar to gesture patterns used in our game, and therefore, their results can also be applied to our game. This means that through certain user training, high recognition rates can be achieved. Although it is not a perfect system, it is sufficient for the game to be operational since it takes advantage of imperfect recognition as a source of challenge. This user training process is exactly what players in our game would strive for – to match the reference gestures through practice.

2.2.3 Speech Recognition

A typical speech recognition system consists of a language model, dictionary and acoustic model. [12] The language model and dictionary are responsible for splitting up words into their base senones (sub-words) after which the pronunciation of each senone is stored in the acoustic model. There are three possible acoustic models: discrete, semi-continuous and continuous hidden Markov model (HMM). The discrete HMM is the fastest but least accurate [13]. The continuous HMM is the most accurate, but at the sacrifice of recognition speed, as it is the slowest. [14] Finally, there is the semi-continuous HMM which fits between the two previous HMM models, with good accuracy while maintaining speed.

Sphinx4 [13] is an open source speech recognition system that has an approximately 97% recognition rate [16]. Although Sphinx4 does not have perfect recognition rate, it is more than sufficient for the purpose of our magic game, where the 3% error can be used as randomness to mimic the field of magic.

3 Design and Implementation

3.1 Game Design

We selected 7 sample spells with different levels of difficulty to emulate the 7 years of Hogwarts School in our prototype design. This demo version offers the user a sample experience of the game; a complete game would likely include an interactive storyline.

Two modes of play are provided are the adventure mode and free play mode. The adventure mode is a flow mode proceeding through all seven spells with increasing difficulty. The free play mode allows the user the flexibility to choose from any spell to learn and practice.

A typical spell learning process starts with a clear and short tutorial showing how to pronounce the spell, how to move the wand, and how to synchronize the two. A cue is used to signal the user to start casting the spell and the recognition system will output one of four outcomes (success, marginal success, failure, and no effect) based on the closeness of the user speech and gesture against the pre-trained database. All spells are extracted from the original Harry Potter series by J. K. Rowling [2] [3]. According to our survey, this is more desired than randomly creating spells. Corresponding gestures of the spells are creatively designed due to the lack of description of these gestures in the book.

3.2 System Overview

The implemented system has two main goals: to accurately capture and interpret the spells casted by the user. As the spells targeted by our system consisted strictly of speech and gesture, a sound capturing device and an accelerometer are adequate to satisfy the need. The inputs provided to the two aforementioned devices are interpreted by recognition software to determine the accuracy of each casted spell. Based on the evaluation, a graphical output is sent to display the results to the user. An overview of the complete system is shown in Figure 2 and is discussed in detail below.

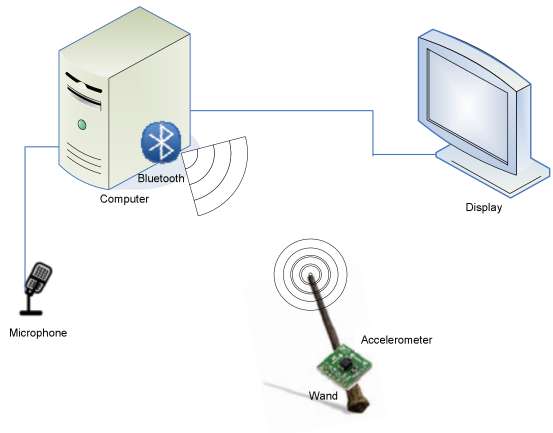


Fig. 2. System overview

3.3 Hardware

Our system consists of a microphone, accelerometer, display and a computer as depicted in Figure 2. For simplicity, we use a laptop containing the microphone, computer, display, and a separate Wii controller with a built-in accelerometer (along with Bluetooth communication components). The Wii controller is used to capture the gestures performed by the user and the microphone is responsible for capturing the speech. The Wii controller and the computer communicate through a Bluetooth connection. A C++ driver created by gl.tter [15] was utilized to facilitate the transmission of data between the Wii controller and the computer. Acceleration triplets, one for each axis{x,y,z}, are used to represent the motion performed by the Wii Controller. The laptop ran Windows 7 with support for C++ and Java programs. In addition, a built-in Bluetooth adapter is included to establish the connection with the Wii remote controller.

3.4 Software

This section discusses the interpretation software employed and the method used to synchronize between the speech and gesture recognition systems.

3.4.1 Speech Recognition

Sphinx-4 developed by CMU [16] was used as the speech interpreter, and was written in Java. This software was chosen because it contains a large vocabulary database, is speaker independent, and has a continuous speech recognition function. Sphinx-4 is based on semi-continuous Hidden-Markov-Model, which matches the input with the most probabilistic word in the database. Our system makes use of the pre-trained Wall Street Journal acoustic models and a custom dictionary was created for each word to indicate the phoneme sequence to pronounce each word.

The speech recognition system saves the matched word and the time which speech was first detected to a text file prior to termination.

3.4.2 Gesture Tracking System

The gesture recognition system is written in C++ and built on top of the Wii remote [15] library which was also written in C++. The Wii remote is sampled at 5 Hz and the result is compared against reference data trained in advance. As gestures vary in length and speed, the total number of data points for each move also varies. This number is fixed during training stage. Sampled user data is clipped from starting time when speech was first detected to the end of gesture length. Comparison is done by calculating error using the Euclidean distance between the two sets of data. The overall error is the arithmetic average of the entire set of data representing the respective gesture.

3.4.3 Integration and Synchronization

The integration between gesture and speech recognition systems was done through synchronized writes and reads from the same text file to allow for communication between the C++ and the Java program. The speech recognition system writes the results to a text file which is read by the C++ program at a later time.

Synchronization information is retrieved by tracking the start time of gesture and speech using the operating system's clock in units of milliseconds. Once the start of speech time is determined, the gesture recognition will search through the accelerometer data and set the closest matching data time as start and proceed with its recognition. Once the start time is determined, the speed of gesture is guaranteed since sampling rate and the number of data points evaluated is constant and the speed of speech recognition system is constant, creating a synchronized overall system. In order to cope with minor system error and human reaction time, an experimentally determined offset of up to 600 ms is accepted between the gesture and speech recognition start time.

3.4.4 Visual Display

There are four possible visual outputs for each spell casted depending on the result of the recognition software. Table 1 presents the visual output possibilities.

Table 1. Visual output decision chart

Speech correct?	Gesture error:	Visual output
Yes	< 0.6	Success
	0.6 – 1.2	Random between shrunk effect and no effect
	> 1.2	Fail
No	< 0.6	Random between shrunk effect and no effect
	0.6 – 1.2	
	> 1.2	Fail

Each spell comes with a corresponding tutorial and each tutorial is an animation consisting of spell pronunciation, wand trajectory, and temporal relation of phonemes in the voiced commands. Figure 3 shows a non-animated picture of the tutorial. The solid black line indicates the wand's path. Phonemes along this line indicate desired synchronization between speech and gesture. After the tutorial, a real scenario would appear as a cue to signal the player to start casting the spell. In the case of 'accio' shown in Figure 4, the scenario is a toad to be magnified, and a successfully casted spell will be an enlarging toad as shown in Figure 6. All animations are made in Adobe Flash CS4 and the GUI is written with Visual Basic. The graphics are quite different from a real video game but is enough for constructing a primitive framework of the entire design.

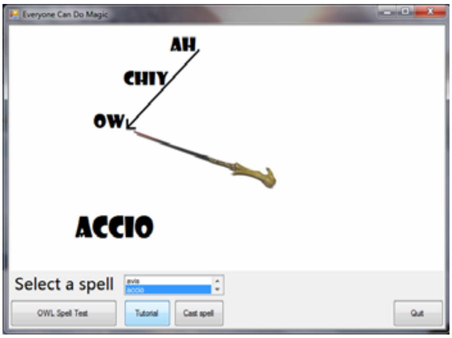


Fig. 3. Sample tutorial

Fig. 4. Cue for user to being casting

Fig. 5. Failed Spell Cast

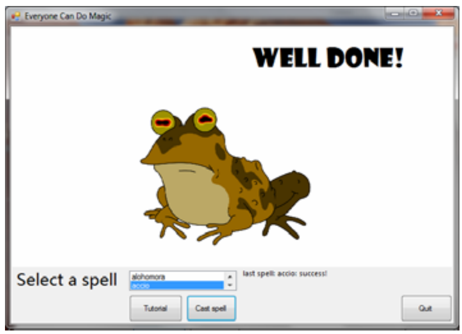


Fig. 6. Success Spell Cast

4 User Study

We adopted a common approach: play testing to evaluate our game design. Twenty-nine subjects (17 males, 12 females) were recruited to participate in the volunteer study (UBC ethics code H07-03063). The principle dependent variable is whether the user enjoyed the game while the independent variable is percentage of spells cast that were successful based on adjusting decision thresholds to make it harder or easier to cast successfully. We call this our challenge level.

4.1 Test Procedure

The user survey was conducted in three different groups based on the success rate of all spells, which were categorized as low, medium, and high. The procedure for each group was the same.

The experimental procedure was first explained to the survey participant. The game then started with a practice mode allowing subjects to manually choose from a list of spells and practice each one at most three times. We allowed players to proceed and see all spells even if they did not pass in the three trials to control the experiment time.

After the user exhausts the maximum number of tries for each spell or upon request, the game will enter the test mode, which loops through all seven spells and records the subject's performance. Our challenge level was adjusted randomly to control the overall success rate in the testing mode for different subjects. Each participant was required to fill out a questionnaire regarding to their experience at the end. We note the number of spells passed on the paper for result analysis purposes. In particular, each subject was inquired about any suggestion for improvement they have, as well as their impressions on the game. For this, they were given choices of anxiety, enjoyment or boredom to describe their feelings through the game.

4.2 Experimental Results

Figure 7 shows the distribution of total performance in test mode for all subjects incorporating their emotions. It can be seen that the highest percentage of each group of participants that felt the game was enjoyable were participants who achieved success rate between 30%-84%. 67% of participants in the high success group were bored, and there was an approximate even distribution of emotions from the subjects in the low success group.

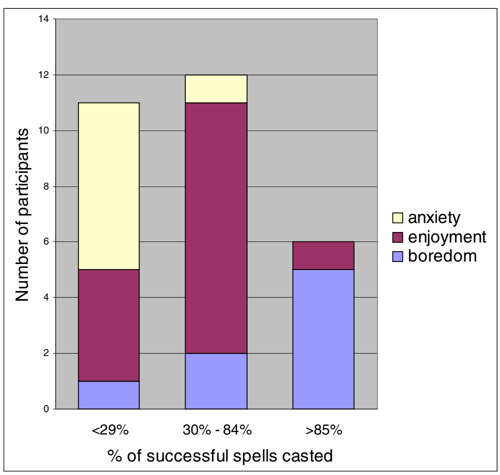


Fig. 7. Game enjoyment vs spell success rate

Figure 8 shows the user enjoyment level of the game with respect to whether the participant is a frequent gamer or not. It can be seen that only 21% (3 out of 14) of frequent gamers didn't enjoy the game, whereas 53% (8 out of 15) of the non-gamers did not enjoy the game.

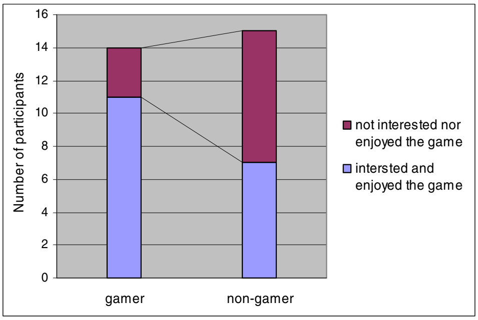


Fig. 8. Game enjoyment vs participant's gaming background

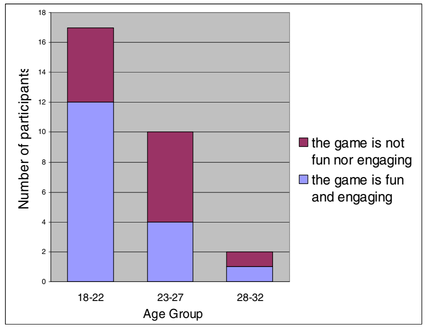


Fig. 9. User feeling vs age

Figure 9 shows user engagement in the game with respect to the participant's age. The majority (71%) of participants in the age group between 18 and 22 years old felt the game was fun and engaging and 40% of the participants in the age group between 23 and 27 years old gave positive feedback. An even split was seen amongst the third group which includes participants between 28 and 32 years, however, this data is inconclusive since the sample size is too small.

4.3 Discussion

Results from Figure 7 suggest that a success rate between 30%-84% provides a significant chance of flow experience where players are most likely to be intrigued by the game. On the other hand, excessive success rates depreciate the enjoyment of the game, which leads to boredom. The result for the low success rate group suggests that participants can still enjoy the game by aggressively improving their skills and striving into the "flow" region despite falling transiently into the anxiety zone.

Figure 7 also demonstrates the correlation between challenge level and user experience, thus, verifying the hypothesis of the three channel model of flow presented in Figure 3 is applicable to our design. One way to obtain the desired success rate on users is to setup an adaptive system where the evaluation criterion becomes more stringent as the player's success percentage exceeds a particular threshold, such as 84%. This way, all players, regardless of skill level can feel challenged without changing the spells. On the contrary, the evaluation criterion could also become lenient for players with a success rate below a specified threshold, such as 30%. The adjustment of evaluation criterion in the background should not be detected by the players, and this adaptive system could be used to maintain the engagement level of all players. Naturally, more complex magical spells can be introduced to provide increasing complexity for advanced skill levels.

By observing the subjects' performance, we noticed that spells with simple speech and gesture are easier than those that consist of a relatively long speech with multiple phonemes and a more intricate gesture motion path, which confirms our preliminary assumptions on spell difficulty.

Results from Figure 8 infer that frequent gamers were much more engaged to the game than non-frequent gamers. Given that frequent gamers are likely to be exposed to various types of games, their past experience establishes a guide for comparing our game against the ones previously played. Positive feedback from them demonstrates that our game has great promise for future investments.

Based on the statistics from Figure 9, our game was more appealing to the age group between 18 and 22 than older age groups. Due to this trend, as well as the target age group for many popular magical paraphernalia, we believe there is likely to be an increasing trend on even younger populations, specifically between ages 13-17 who would be attracted by the game.

Overall, the subjects were fascinated by the game despite that 6 out of 29 testers thought the game's visual interface could be improved, which encourages further investigation on this type of game. The novelty of the design idea and integrated system with both speech and gesture recognition technology explores a new space for future game development. Many subjects recommended that a better plot, background music and costumes such as a pointy hat for the players would further improve game quality.

5 Conclusion

In this paper, we presented the design of an interactive game using speech and gesture recognition systems. We created a framework with a spell casting environment which took advantage of the imperfection of the aforementioned technology to portray randomness in the magical world. The user study showed that frequent gamers aged under 22 could be the main target audience for such a game. In order to maintain the player's interest, the design should follow the principles of the three channel flow model and the spell casting success rate should be kept within 30% and 84%. This success rate would avoid leaving players with feelings of anxiety or boredom.

6 Future Work

Current data suggests that most players were only able to successfully cast 2 or 3 spells out of the 7 with minimal practice. As we plan to introduce a more complete story plot, more spells are needed.

Due to ethical issues involved with running experiments on minor subjects, we were unable to obtain feedback from users in younger age groups. As seen from Figure 10, there is a clear trend of increasing interest for our game for younger participants. Based on this, more experiments on younger subjects could be done to obtain important user feedback.

Despite a suboptimal storyline, enhancing visual interface with more vivid animations and completing the storyline with more spells and versatile activities are the main future work of this design. Some details such as flexibility of skipping part of the introduction and tutorial more help information, and better wand design could also help make it more like a commercial game. Furthermore, user customizations such as selecting courses according to user interest and picking their own wands can be added to furnish the design.

Another aspect that is missing in our prototype system is providing audio clips for visual animations. A door that opens with a squeaking sound is much more satisfying than only seeing the door open on the visual output. The user would thus experience a more realistic environment and surely be more immersed in the game.

Unifying the programming language for speech and gesture recognition software can further improve the synchronization and the integration of the overall system. One possibility is to rewrite either the speech recognition in C++ or rewriting the Wii remote driver in Java. Either way, using one language would make it easier to access intermediate recognition information by the other system, allowing better synchronization schemes.

Acknowledgements

This work has been partially supported by The Institute for Computing, Information, and Cognitive Systems (ICICS) at University of British Columbia.

References

1. Pinelle, D., Wang, N., Stach, T.: Heuristic Evaluation for Games: Usability Principles for Video Game Design. In: CHI 2008, vol. 5, pp. 1453–1462 (April 2008)
2. Rowling, J.K.: Harry Potter and the Philosopher's stone (1997)
3. Rowling, J.K.: Harry Potter and the Goblet of Fire (2000)
4. Lecture 12: An overview of Speech Recognition, http://www.cs.rochester.edu/u/james/CSC248/Lec12.pdf (retrieved February 25, 2010)
5. Kiili, K.: Content creation challenges and flow experience in educational games: The IT-Emperor case. The Internet and Higher Education 8(3), 183–198 (2005)
6. http://www.xbox.com/en-US/kinect/default.htm (retrieved June 18, 2010)
7. http://en.wikipedia.org/wiki/Lego_Harry_Potter (retrieved March 12, 2010)
8. http://www.ea.com/games/harry-potter-half-blood-prince (retrieved March 12, 2010)
9. Buxton, B.: Chapter 14: Gesture Based Interaction (May 2009), http://www.billbuxton.com/input14.Gesture.pdf (Retrieved March 6, 2010)
10. Leong, T., Lai, J., Pong, P., et al.: Wii Want to Write: An Accelerometer Based Gesture Recognition System. In: International Conference on Recent and Emerging Advanced Technologies in Engineering 2009, Malaysia (2009)
11. Mlích, J.: Wiimote Gesture Recognition. In: Proceedings of the 15th Conference and Competition STUDENT EEICT 2009, vol. 4, pp. 344–349 (2009)
12. Q6.1: What is speech recognition?, http://www.speech.cs.cmu.edu/comp.speech/Section6/Q6.1.html (retrieved February 25, 2010)
13. Walker, W., et al.: Sphinx-4: A flexible open source framework for speech recognition (2004), http://research.sun.com/techrep/2004/smli_tr-2004-139.pdf
14. Huang, X., Alleva, F., Hon, H.W., Hwang, M.Y., Rosenfeld, R.: The SPHINX-II speech recognition system: an overview. Computer Speech and Language 7(2), 137–148 (1993)
15. http://wiiyourself.gl.tter.org/ (retrieved March 16, 2010)
16. http://cmusphinx.sourceforge.net/sphinx4/ (retrieved March 18, 2010)

Onomatopen: Painting Using Onomatopoeia

Keisuke Kambara and Koji Tsukada

Ochanomizu University, Ochadai Academic Production,
2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan
Tel.: +81-3-5978-2571
kambara@sappari.org, tsuka@mobiquitous.com
http://sappari.org/

Abstract. We propose an interactive technique using onomatopoeia. Onomatopoeia are imitative words such as “Zig-zag” and “Tick-tock”. Some Asian languages, especially Japanese and Korean, have many onomatopoeia words, which are frequently used in ordinary conversation, as well as in the written language. Almost all onomatopoeic words represent the texture of materials, the state of things and emotions. We consider that onomatopoeia allows users to effectively communicate sensory information to a computer. We developed a prototype painting system called Onomatopen, which enables a user to switch brushes and apply effects using onomatopoeia. For example, if the user draws a line while saying “Zig-zag Zig-zag...”, a jagged line is drawn. As a result of our user test, we found that users can easily understand the usage and enjoy drawing with this application more than with conventional painting software.

Keywords: Onomatopoeia, Multimodal input, Pen-based interface, Voice-based interface.

1 Introduction

Onomatopoeia is a general expression used in ordinary spoken and written language. Some kinds of onomatopoeic words imitate sounds, such as the sound of a clock: “Tick-tock”. Others mimic states or emotions, such as “Zig-zag”.

The use of onomatopoeia varies with language, and some Asian languages, especially Japanese and Korean, have many onomatopoeia words.

Onomatopoeic words represent states, movements, feelings and emotions, and allow their expression in a fun, lively manner. Children typically call cars “Zooom” or call dogs “Bow-wow”, as onomatopoeic words are easy to pronounce and remember even for children.

Although onomatopoeic words are often used in conversation and writing, they are not used in human-computer interactions. Among the many research projects on interactive techniques using voice input, few research projects have focused on onomatopoeia. Herein, We propose a friendly, sensory-based interactive technique based on the use of onomatopoeia.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 43–54, 2010.

2 Onomatopen

We applied the interactive technique using onomatopoeia to painting software, and developed a prototype called Onomatopen. Onomatopen enables users to draw textured lines using onomatopoeia, and the texture is associated with the onomatopoeic word spoken. In case of drawing a dotted line, the user draws a line while saying "Ten Ten...[1]", and the texture of the line becomes dotted (Fig. 1). In a similar way, using various onomatopoeia, the user can control certain functions of painting software such as switching brushes and applying effects to images.

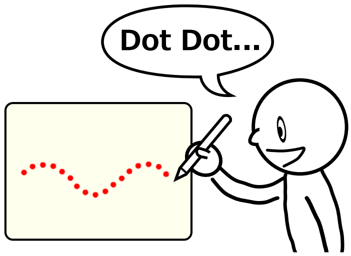


Fig. 1. Concept Image of Onomatopen

2.1 Features

The novel features of Onomatopen compared with existing painting software are as follows.

- Suitable mapping between operation methods and functions
- Multimodal interaction
- Entertainment applications

Suitable Mapping between Operation Methods and Functions. Onomatopoeia is suitable for representing states of objects in a more sensory-based manner. In Onomatopen, users can imagine the texture of a line or executed function more easily than by changing brushes using buttons.

It is important to associate operation methods with functions intuitively in painting software, because drawing pictures requires sensitivity and feeling.

Multimodal Interaction. One of the important features of Onomatopen is the ability to draw lines using the voice. This is a kind of multimodal interface. We can help users input more details, such as texture, while drawing. Furthermore, this interaction omits explicit switching operations such as selecting a brush.

Entertainment Applications. The above features of Onomatopen, such as appealing to sensitivity and visual/auditory feedback, are suited to the entertainment of the user. Users enjoy synchronizing the pen input and visual/auditory feedback. Drawing for play or as a hobby needs to be fun not only efficient.

[1] "Ten" means a dot in Japanese.

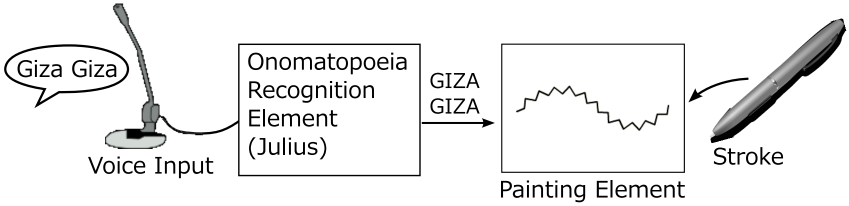


Fig. 2. System architecture

2.2 Functions of Onomatopen

We have currently implemented 10 features using onomatopoeia (Table 1, Fig. 3). There are two main uses of onomatopoeia; switching brushes and editing images.

Table 1. Functions corresponding to onomatopoeia

Onomatopoeia	Function (Edit, Effect)
Keshi Keshi (means erasing)	Eraser
Choki Choki (means cutting)	Cut
Peta Peta (means pasting)	Paste
Copy Copy	Copy
Biyoon (means deformation)	Deformation

Switching Brushes. In the case of general painting software, brushes are switched using menus or keyboard shortcuts. On the other hand, users can change brushes simply by saying the appropriate onomatopoeia for the desired texture in Onomatopen (Fig. 1). For example, by drawing while repeatedly saying "Zig-zag Zig-zag...", the texture of the brush automatically switches to Zig-zag. Figure 3 shows the brushes available and the corresponding onomatopoeia.

Editing Images. We have also developed editing functions such as erasing and cut/paste using onomatopoeia.

Users can erase drawn pictures by stroking the picture while saying "Keshi Keshi...[2]" (Fig. 4).

We have also developed copy/cut/paste functions using onomatopoeia. Pictures are cut by stroking while saying "Choki Choki...[3]" and copied by stroking with saying "Copy Copy". Cut or copied pictures are pasted by clicking while saying "Peta[4]". By stroking with saying "Peta Peta...", the picture is pasted repeatedly and acts like a kind of brush (Fig. 5).

Deformation Effect. We have developed a function to apply a deformation effect using onomatopoeia. By dragging a picture while saying "Biyoon[5]", the dragged area is stretched like rubber (Fig. 6).

[2] "Keshi" means erase in Japanese.
[3] "Choki" means snip in Japanese.
[4] "Peta" means pasting in Japanese.
[5] "Biyoon" represents the sound of stretching a rubber-like material in Japanese.

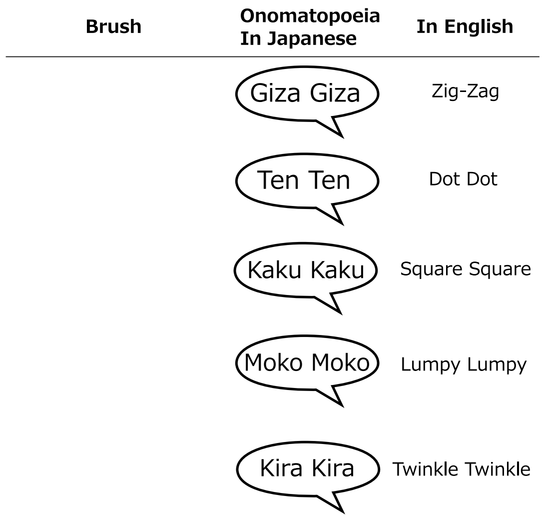


Fig. 3. Brushes corresponding to each onomatopoeia

Fig. 4. Erase function

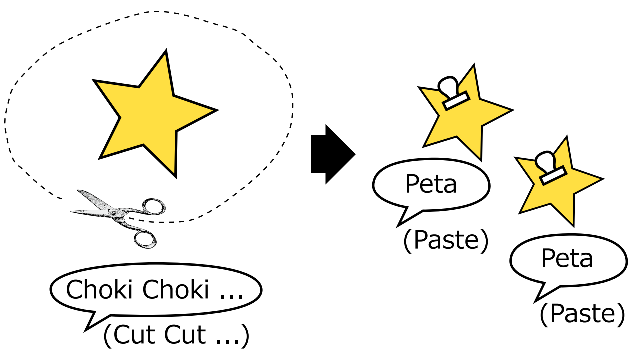


Fig. 5. Cut & Paste

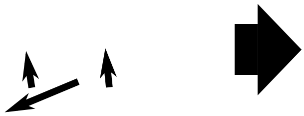


Fig. 6. Deformation Effect

3 Implementation

The Onomatopen system consists of two parts; the voice recognition element and the painting element (Fig.2). The voice recognition element extracts onomatopoeia from user's voice in real-time. The painting element receives the results from the recognition element and switches to the function corresponding to the onomatopoeia.

3.1 Voice Recognition Element

The voice recognition element is based on Julius[1], which is an open-source voice recognition engine. Julius allows the user to customize the grammar for recognition. Onomatopen uses the grammar that recognizes repeated onomatopoeia. Figure 7 show the state transition diagram for the recognition of the repeating onomatopoeia "Zig-zag Zig-zag...".

Julius runs as a socket server, which sends the result from the recognition element to the painting element immediately after recognition, the painting element requires the recognition results while the pen is in motion and the onomatopoeic phrase is being repeated.

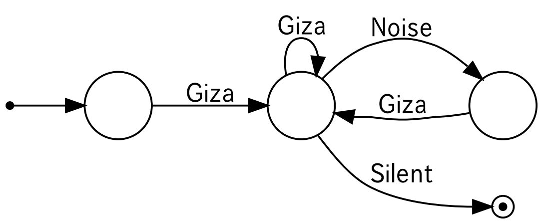


Fig. 7. State transition diagram recognizing "Giza Giza..."

3.2 Painting Element

The painting element was implemented in Adobe AIR. This element receives the onomatopoeic words from the voice recognition server and switches to the brush corresponding to the onomatopoeia.

It is a burden for the user to say the onomatopoeic phrase repeatedly throughout stroking. To reduce the burden, even after the user stops repeating the phrase, the brush continues as selected until the end of the stroke. The painting element corrects voice recognition errors as described below.

3.3 Voice Recognition Error Correction

The voice recognition system cannot avoid recognition error. If the brush reflected the recognition result without correction, it would frequently switch to an unintended format.

To correct recognition error, we took advantage of the following characteristic of onomatopoeia. In Japanese, almost all onomatopoeia consists of the repetition of the same word. Conversely, it is unusual for other words to be added in repetitions; for example, “Zig-zag Dot Zig-zag...”. When such an unusual pattern is recognized, the onomatopoeia pattern is corrected as a repetition of “Zig-zag”.

Using this method, the brush changes format on the basis of the most frequently appearing word. When the brush changes format during a stroke, the drawn line is rendered again from the start point.

According to this correction method, if an unintended line is drawn, the user can fix it by simply repeating the correct word several times (Fig.8).

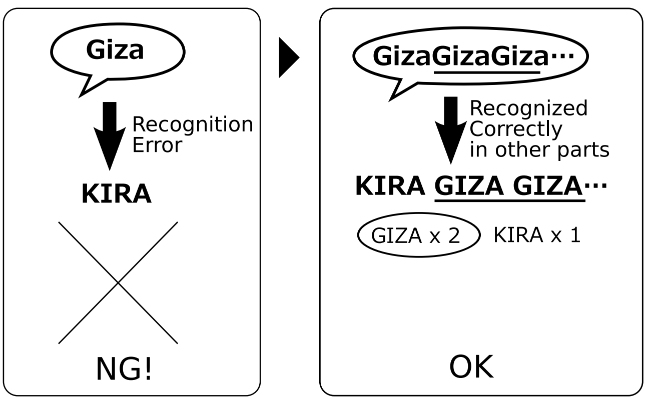


Fig. 8. Error correction by repeating onomatopoeia

3.4 Gap between the Beginning of a Stroke and the Beginning of Speech

Oviatt[10] showed that the order and the gap between the beginning of a stroke and the beginning of speech differs according to the individual. If the user occasionally begins repeating the onomatopoeia after the beginning of the stroke, the type of brush can not be defined from the beginning. In such case, a light gray line is drawn at the beginning, and this line is drawn again after the brush format has been defined by voice recognition. The user can draw lines regardless of the gap between the beginning of a stroke and the beginning of speech.

4 Evaluation

We evaluated the prototype by user testing and observation.

4.1 Experiment

After brief instruction regarding the features, each subject freely drew pictures using the Onomatopen for about 10 minutes (Fig.9). Last, we surveyed the subjects by a questionnaire containing questions scored from 1 to 5 and by a free form.

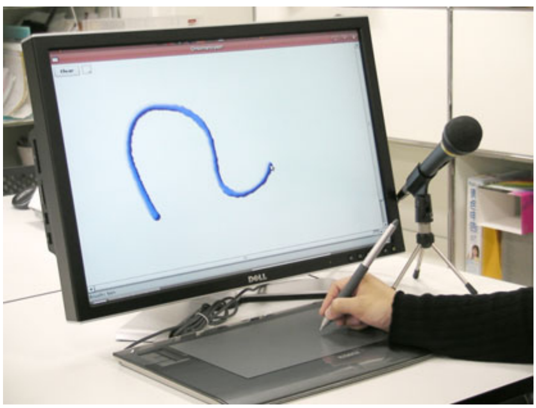


Fig. 9. Evaluation Setting

4.2 Subjects

9 subjects (22-32 years old Japanese, 7 females and 2 males) participated in this experiment. The study group contained an expert in painting who uses painting software almost everyday and 3 people who had had little experience with painting software.

4.3 Results and Discussion

Tables 2-5 show the results from the questionnaire.

Table 2. Did you feel the mapping between functions and operations natural?

Natural	6 3 0 0 0	Unnatural

Table 3. Could you soon understand the usage?

Yes, immediately	8 1 0 0 0	No, not at all

Correspondence between Functions and Operations. As shown in Table 2, all subjects answered that the mapping between functions and operations was "Very Natural" or "Natural". This result supports the idea that switching the brushes and functions using onomatopoeia is an intuitive interaction.

Eight of 9 subjects could remember the usage immediately (Table 3); therefore, it is reasonable to suppose that the interaction technique is a easy to remember regardless of one's experience with painting software.

One of the reasons why subjects could remember all the onomatopoeia easily is that the number of onomatopoeic phrases recognized in the current prototype is limited in 10. If the number of onomatopoeic phrases is increased, it will be hard to remember those onomatopoeic words. Although this prototype implements brushes and functions that are easy to express in onomatopoeia, some of brushes and functions have no/many corresponding onomatopoeia.

Table 4. Could you draw as you like without failure?

Yes, perfectly	1 2 2 4 0	No, not at all

Table 5. Could you enjoy drawing?

Enjoyed	7 0 1 1 0	Bored

Textures or functions imagined from onomatopoeic phrases differ according to the individual. In case of the experiment, a subject commented that "The lumpy texture is different from what I imagined". Thus it became clear that brushes implemented in this prototype did not always correspond to what the subjects imagined.

For this reason, the following points require further survey; the kind of brush associated with certain onomatopoeia, the number of onomatopoeia that a user can remember, and individual differences in the understanding of specific from onomatopoeia.

Although it is unrealistic to control all functions in existing painting software using onomatopoeia for the above-mentioned reasons, we suppose it is feasible to assign some functions to specific onomatopoeia as shortcuts. In addition to the normal controls using GUI, by assigning onomatopoeia to particular functions which are frequently used or are easy to express in onomatopoeia, the user will not have to remember all onomatopoeia, and be gradually able to learn the necessary onomatopoeia. If the user can add/change the assignment, individual differences will not be a problem.

Multimodal Interaction. The input method of Onomatopen, which uses a pen and voice simultaneously, is not popular in current human-computer interactions. Although all subjects could understand the basic usage immediately, 4 of 9 subjects answered they could not draw as much as they liked (Table 4), and the prototype received low scores. One subject commented that “Pasting is a little difficult to use. Occasionally voice recognition failed, and it did not work correctly.”. Since the onomatopoeia corresponding to the stamp tool and the deformation effect are not pronounced repeatedly, error correction based on onomatopoeia repetition did not work, and failures often occurred when there was a gap between the beginning of the stroke and speech. Another reason for the low evaluation is the lack of an Undo/Redo function. This prototype did not have an Undo/Redo function and the user could not undo a failed stroke. The problem with the user not being able to draw as much as they liked requires further work and we hope it will be resolved by the implementation of Undo/Redo and the reduction of voice recognition errors.

Error correction by onomatopoeia repetition worked. Subjects who understood the function could draw lines as intended by repeating the onomatopoeic word when the line was incorrect at the beginning of the stroke. Although subjects did not notice the function in the beginning of the experiment, they could quickly understand its use after it was explained and they were able to employ the function.

We plan to carry out quantitative evaluation of the error correction function and detailed analysis of the gap between the stroke and speech in order to improve the multimodal interaction.

Entertainment. Seven of 9 subjects felt that drawing using Onomatopen was fun (Table 5), and 5 subjects commented that it was very enjoyable. It is likely that Onomatopen entertains the user more than existing painting software, although there were subjects who could not draw as much as they liked. On the other hand, some subjects commented that “It’s also fun to fail” and “It will be exciting if I draw with many people”. If Onomatopen is used for entertainment, the failures will be simply a part of playing.

5 Possible Applications

Further possible applications of Onomatopen include the following.

5.1 Maps

One merit of Onomatopen is that the user can draw decorated lines quickly and easily. When a user draws a map for guidance, it is better that the lines representing rails, streets and rivers are shown differently. For example, rails and trains are drawn by saying "Clickety-clack", streets and cars are drawn by saying "Zoom". In this way, we believe the user can draw maps more efficiently.

5.2 Animation and Games

It may be possible to move still pictures using onomatopoeia representing movement. An airplane can be made to fly by saying "Zoom" or a train made to run by saying "Clickety-clack".

In the similar way, video games can be controlled by onomatopoeia. For example users can make characters jump by saying "Hop!" or fire a gun by saying "Bang!".

6 Related Works

6.1 Multimodal Interface Using Voice Input

VoicePen[4] proposed a technique to control the thickness, color, transparency and definition of lines using voice characteristics such as loudness and pitch. Although it is similar to Onomatopen in that the textures of lines is controlled by the voice, our research focused on onomatopoeia as information provided by the voice.

Igarashi et al.[5] and Harada et al.[3] proposed interactive techniques using non-verbal information, and reported that non-verbal input enables users to continuously control the strength of effects with resistivity against noises. It will be possible that Onomatopen can adopt such a technique for controlling the thickness and transparency of lines by the volume or pitch of the voice, in addition to the control by onomatopoeia alone.

A multimodal interface that combines pen input and voice commands has been discussed for a long time[2][6]. Onomatopen is different in that it is focused on onomatopoeia for voice input.

Kurihara et al.[8] proposed a system that supports writing on a electronic blackboard during lectures by predictions, and allowing the completion of handwritten notes from the talk. They reported that it was designed to aid writing more quickly, easily and clearly to record information provided in lectures. Onomatopen will also be able to support such situations.

6.2 Sound Symbolism of Onomatopoeia

Onomatopoeia is mainly studied in the fields of literature, linguistics and psychology. A representative study is Tamoris' investigation of the phonological characteristics of onomatopoeia [11]. Murakami[9] analyzed the sound symbolism, which is the relation between the phoneme and the meaning. He extracted phonemic components from Japanese onomatopoeic words by multivariate analysis, and clarified the relation between the phoneme and the meaning in terms of emotions, impressions and materials.

In the field of computer science, Komatsu demonstrated technology to digitize the impressions from onomatopoeia based on the acoustic meaning of the phoneme [7]. Each onomatopoeia is expressed by 8 attribute values such as hardness, smoothness and humidity. We may apply this technology to automatically associate an onomatopoeic word with a brush. Another approach to generate new brushes could be also possible through the analysis of sounds and directly extracting the symbol for the sound without recognizing the spoken word. While the former approach uses only symbolized and simplified information from the sound, the latter approach has an advantage in using much more information from the sound. With the latter approach it may be possible to control the expression based on the way of pronunciation. However, loudness and pitch differ according to the individual, thus the general ability that anybody can draw the same zig-zag line by saying "Zig-zag" would be lost.

7 Conclusion

We developed a novel drawing system, called Onomatopen, that enables users to draw various textured lines by saying aloud the corresponding onomatopoeia. In addition to drawing lines, users can edit and apply effects to images using onomatopoeia. This input technique enables users to draw in a more sensory-based manner and more enjoyable than with existing painting software. We will apply our interactive technique to other areas in which we can takes advantage of onomatopoeia including animation, games and drawing maps.

References

1. Grammar-based Continuous Speech Recognition Parser: Julius, http://julius.sourceforge.jp/
2. Cohen, P.R., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L., Clow, J.: Quickset: multimodal interaction for simulation set-up and control. In: Proceedings of the fifth conference on Applied natural language processing, pp. 20–24 (1997)
3. Harada, S., Landay, J.A., Malkin, J., Li, X., Bilmes, J.A.: The vocal joystick:: Evaluation of voice-based cursor control techniques. In: Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 197–204 (2006)
4. Harada, S., Saponas, T.S., Landay, J.A.: Voicepen: Augmenting pen input with simultaneous non-linguistic vocalization. In: Proceedings of the 9th International Conference on Multimodal Interfaces, pp. 178–185 (2007)
5. Igarashi, T., Hughes, J.F.: Voice as sound: Using non-verbal voice input for interactive control. In: Workshop on Interactive Systems and Software IX, pp. 7–10. Japan Society for Software Science and Technology (2001)
6. Julia, L., Faure, C.: Pattern recognition and beautification for a pen based interface. In: Proceedings of the Third International Conference on Document Analysis and Recognition, vol. 1, pp. 58–63 (1995)

7. Komatsu, T., Hinata, K.: Onomatopoeic filtering system for assisting user's intuitive expressions. In: The 22nd Annual Conference of the Japanese Society for Artificial Intelligence, pp. 1D2–10 (2008)
8. Kurihara, K., Gogo, M., Ogata, J., Igarashi, T.: Speechpen: Predictive handwriting based on ambient multimodal recognition. In: Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 851–860 (2006)
9. Murakami, Y.: The inquiry of phonetic symbolism hypothesis: Analyses of onomatopoeias by the clustering of the introduced components no phonemes, sd technique, associated nouns, and associated verbs. The Japanese journal of educational psychology 28(3), 183–191 (1980)
10. Oviatt, S., Lunsford, R., Coulston, R.: Individual diffrences in multi-modal integration patterns: what are they and why do they exist? In: CHI 2005, pp. 241–249 (2005)
11. Tamori, I., Kakehi, H.: Paradise of Onomatopoeia, Echoic word, Imitative word. Keiso shobo (1993)

Helping Hands: Designing Video Games with Interpersonal Touch Interaction

Cody Watts[1,2], Ehud Sharlin[1], and Peter Woytiuk[2]

[1] 1 University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada
[2] BioWare Corp., Edmonton, Alberta, Canada
{wattsc,ehud}@cpsc.ucalgary.ca, woytiuk@bioware.com

Abstract. Increasingly, the movements of players' physical bodies are being used as a method of controlling and playing video games. This trend is evidenced by the recent development of interpersonal touch-based games; multi-player games which players control by physically touching their partners. Although a small number of interpersonal touch-based games have recently been designed, the best practices for creating video games based on this unconventional interaction technique remain poorly explored and understood. In this paper, we provide an overview of interpersonal touch interaction in video games and present a set of design heuristics for the effective use of interpersonal touch interaction in video games. We then use these heuristics to analyze three current interpersonal touch-based games in order to show how these heuristics reflect on the current state of the art. Finally, we present our vision for the future of this interaction modality in video games.

Keywords: Games, interpersonal, touch, interface, design, heuristics.

1 Introduction

Within the last decade, the way we play video games has changed. The phenomenal success of games like *Dance Dance Revolution, Guitar Hero* and *Rock Band* has ushered in a new era of physically-active gaming. To see the popularity of these physically-active games, one need look no farther than Nintendo's home fitness game, *Wii Fit* which (as of this writing) has sold 22.6 million copies, making it the fifth-best-selling video game ever produced. The success of these games and of many others like them has forced both players and game designers to broaden their understanding of what it means to play video games. Video gaming is no longer just an activity which involves passively staring at a computer monitor – it can also be an active experience, one which involves the player's entire body.

One of the recent advancements in the field of physically-active gameplay is the advent of video games based on interpersonal touch. These are multiplayer video games in which the system can sense when two or more players are physically touching, and which makes the act of touch part of the game. The distinguishing characteristic of interpersonal touch-based games is not just that the players' physical actions have a direct impact on the game they are playing, but that these actions arise through their interactions with their fellow players. In this regard, games based on

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 55–66, 2010.

interpersonal touch provide an inherently social experience, since their gameplay necessitates interaction between players.

We are aware of only four video games which have made use of interpersonal touch in their gameplay mechanics: *"Get Lucky" Charms* (also known as *Intimate Controllers*) by Chowdhury [3], *Freqtric Game* by Baba [1], and *Matchmaker* and *Prism Squad: GO!*, both designed by our group [11, 10, respectively]. As these four games form the basis of our discussion in this paper, we shall examine each of them in detail later on. But first, we turn our attention to a more fundamental question: How does interpersonal touch fit into the broader practice of video game design?

2 Interpersonal Touch in Video Games

If we accept that the purpose of a game is to provide amusement to its players, then any effective use of interpersonal touch must somehow contribute to this overarching goal. We believe that interpersonal touch can best contribute to players' enjoyment by promoting player-to-player socialization. The value of socializing in multiplayer video games should not be underestimated; several studies have shown that the social atmosphere created by video games can be just as important to players' enjoyment as the game they are playing. In "Why We Play Games: Four Keys to More Emotion Without Story", Lazarro examined thirty volunteers while they played their favorite video games in an attempt to understand what makes games fun [8]. She encapsulated her findings in the form of four "keys" – aspects of gameplay which players enjoyed. Lazzaro's fourth key, which she called "The People Factor," described the rewarding interactions which come from socializing with other players. Lazarro observed that "players in groups emote more frequently and with more intensity than those who play on their own. Group play adds new behaviors, rituals, and emotions that make games more exciting." In fact, her findings indicate that the act of socializing may often take precedence over the game itself – in many cases, players would "play games they don't like [just] so they can spend time with their friends." [8]

Voida et al. reported similar findings in their study of collocated gaming groups [9]. In their own words: "The primary motivation for group console gaming was not the games themselves, but the social interactions afforded by the collocated gameplay. The most important part of group console gaming was, very simply, 'the sociability of it'". In fact, for many participants in Voida's study, socialization was not just the primary motivation for playing but the only reason: many of the adult females and all of the elderly participants who were interviewed for the study admitted that they only played console games in groups – never by themselves [9].

We believe that interpersonal touch provides opportunities for socialization in a media where social interaction is desirable. Human beings naturally recognize touch as a gesture which signifies a meaningful social connection; when two players touch, it acts as an icebreaker and as an invitation to further social interaction – even if the players themselves don't consciously realize this [4]. By integrating interpersonal touch into the designs of video games, we believe it becomes possible to create more social, more memorable gameplay experiences.

3 Video Games Based on Interpersonal Touch

With our newfound understanding of interpersonal touch and how it shapes players' gameplay experiences, let us now examine how interpersonal touch has been used in practice. We begin our examination of interpersonal touch in games with *Intimate Controllers* [3] – a project which combined video gaming with the intimate aspects of touch.

3.1 Intimate Controllers

Intimate Controllers was an exhibit presented at Unravel, the SIGGRAPH 2007 fashion show. The eponymous intimate controllers were pair of wearable undergarments – a woman's bra, and a man's boxer-short – designed to be used as wearable video game controllers. Each controller contained a set of embedded touch sensors which were divided into three "intimacy levels". The more intimate touch-sensors were positioned nearer to the cups of the bra and the buttocks of the shorts, respectively. These controllers were designed to be used by couples as a way of encouraging intimate interaction between the partners as they played. Each player's inputs are located on the opposite player's body – the male player touches his partner's bra, and the female player touches her partner's underpants. Due to the layout of the sensors, this leads to very sexually-explicit positioning when two partners play together: the male stands behind his female partner, cupping her bra, while the female player reaches behind herself to cup the male's buttocks [3].

Accompanying these *Intimate Controllers* was a video game entitled *"Get Lucky" Charms. "Get Lucky" Charms* was a timing-based game similar to *Dance Dance Revolution* in which players had to touch their partner in the appropriate location as symbols corresponding to the various intimacy levels scrolled from the bottom to the top of the screen.

3.2 Freqtric Game

Freqtric Game is the unifying label applied to three interpersonal touch-based games developed by Baba et al [1]. Each game in the *Freqtric Game* series is played using the "*Freqtric Game* controller device" – a handheld gamepad with a steel plate on the back which enables the system to detect when two players are touching. The three games each explore the use of touch in a different genre of gameplay. *Freqtric Shooting* is a two-player cooperative top-down shoot-'em-up where players can touch their partner to activate screen-clearing bombs. Defeated players can also rejoin the game by touching their partner 100 times in a row. *Freqtric Dance* is a *Dance Dance Revolution* clone which introduces a "touch" symbol to the standard directional arrows. *Freqtric Robot Battle* is a competitive two-player fighting game which plays out like a robot sumo match; each player's goal is simply to knock their opponent out of the ring. In addition to firing a virtual missile launchers, players can attack their opponent's robot by physically attacking their opponent; slapping the other player will cause their robot to recoil, while grabbing your opponent will allow your robot to drag your opponent's robot around the arena.

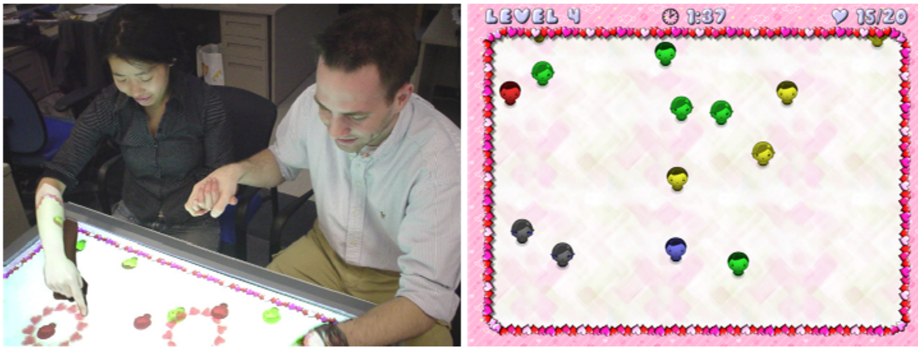


Fig. 1. A couple playing *Matchmaker* (left). *Matchmaker*'s main game screen (right). [11]

3.3 Matchmaker

Matchmaker is a two-player, cooperative tabletop video game which, like *Intimate Controllers*, examines the role of touch as a symbol of romantic love between couples [11]. But whereas *Intimate Controllers* put its emphasis on sexual intimacy, *Matchmaker* focuses on the cute, more innocent aspects of love. [11]

In *Matchmaker*, players work as a team to create matches between tiny virtual people called "Peeps" (Fig. 1). Players can move Peeps around the tabletop by touching and dragging them with their fingers. When the players select two Peeps and drag them together, a match will be created if the Peeps are the same color. The players' goal is simply to make as many matches as they can. However, there is a catch: If Peeps are not matched up quickly enough they will become "lovelorn". Lovelorn Peeps quickly turn grey, making them ineligible to be matched with other Peeps. Lovelorn Peeps can only be cured through interpersonal touch. If a player holds their partner's hand and touches a lovelorn Peep then the Peep's color will return, allowing it to be matched up once more.

Matchmaker shows how interpersonal touch can serve two functions – it is both a gameplay mechanic and a social dynamic. In gameplay terms, handholding is a cooperative mechanic which encourages players to work together in their pursuit of a shared goal. However, it is also a tangible symbol of the game's romantic theme.

3.4 Prism Squad: GO!

Whereas the previous two games explored the use of interpersonal touch in romantic contexts, *Prism Squad: GO!* was our attempt to broaden the appeal of interpersonal touch in games by creating a touch-based game which could be enjoyed by anyone – not just romantic couples [10]. *Prism Squad: GO!* is a science-fiction shoot-'em-up, in which players defend a planet from invading UFOs by using their personal spacecraft to shoot down the incoming UFOs before they can crash into the surface of the planet. The game is played using Nintendo Wiimote controllers and a large-screen display.

Prism Squad: GO! is a game which makes extensive use of color. Both the players and the enemies are color-coded. This is significant because an enemy UFOs can only be destroyed by a player of the same color. And while there are only three players (red, yellow and blue), the enemies come in seven distinct colors (red, yellow, blue, orange, green, purple and white.) While the red, yellow and blue UFOs can be destroyed by a single player acting alone, players must use teamwork to defeat the orange, green, purple and white UFOs. This teamwork is manifested through *Prism Squad*'s color-blending mechanic: whenever two players touch, their colors will temporarily be combined, changing the color of each player involved. For example, if the red and blue players touch, their spacecrafts will each turn purple, temporarily allowing them to destroy purple UFOs. If all three players touch, their spacecrafts will turn white. Much of the excitement in *Prism Squad: GO!* comes from deftly coordinating blends with your teammates to take on the ever-changing onslaught of colorful UFOs.

Although *Prism Squad: GO!*'s gameplay mechanics were inspired by interpersonal touch, we are currently still developing a working touch-sensor for the game [10]. Therefore, we used a form of "simulated touch" to develop and playtest *Prism Squad: GO!* We designated the 'A' button on each player's Wiimote as the "touch button" – whenever two or more players pressed this button simultaneously, the game client behaved as if those players were touching and blended their colors together. Although this method of simulating touch cannot capture the physical sensation of interpersonal touch, we believe it adequately simulates interpersonal touch's cooperative nature.

We evaluated *Prism Squad: GO!* in a series of playtest sessions with volunteer participants [9]. Observations of the players' behavior, combined with testimony from the players themselves suggested that *Prism Squad*'s cooperative, team-based gameplay encouraged players to communicate and interact with their teammates as they played. When the game ended, we asked participants to describe how they thought the game would be different if interpersonal touch was used to blend colors instead of simultaneous button-pressing. Although many participants felt that interpersonal touch would further encourage the rewarding inter-player interactions which players already enjoyed, some participants warned that the addition of interpersonal touch could make it awkward to play the game with strangers, or with members of the opposite sex. Some participants also argued that it may be difficult to get close enough to physically touch your partner, especially in situations where a third player is blocking your way.

4 Heuristics for Designing Games Based on Interpersonal Touch

Like all interaction techniques, interpersonal touch has its own pitfalls, limitations, and best practices. However, because there are so few video games which have made use of interpersonal touch in their gameplay mechanics, there have been very few opportunities to study this interaction technique in practice. This places game designers who are interested in creating video games based on interpersonal touch at a disadvantage; without a foundation of understanding of how to use interpersonal touch effectively, every designer must – through trial and error – devise their own best practices from scratch. The following heuristics are our attempt to address this situation. Here, we present our guidelines for the effective application of interpersonal touch interfaces to video games. These heuristics come directly from the lessons we learned designing, developing, refining and evaluating *Matchmaker* and *Prism Squad: GO!* These heuristics can be used as both a framework to evaluate a video game's use of interpersonal touch (as we will demonstrate in section 5), or as suggested guidelines for the design of future games based on touch. Our heuristics are:

i. Use Touch to Simplify Collaborative Interactions between Players
ii. Use Touch to Create Challenge
iii. Use Touch to Encourage Socialization between Players
iv. Manage Players' Social Expectations
v. Design for the Physical Limitations of Touch

i. Use Touch to Simplify Collaborative Interactions between Players

One of the understated advantages of interpersonal touch is that it is a simple and natural way for two or more players to interact with each other. Touch is never something which needs to be taught or explained. It is simple and natural to reach out and touch someone as a method of demonstrating a connection with that person and it is just as simple to withdraw your hand to break such a connection. In any gameplay situation where two or more players are performing some action which affects them jointly, interpersonal touch is a quick and easy way to link these players for the duration of their action.

ii. Use Touch to Create Challenge

Although interpersonal touch can sometimes streamline interactions between players, touch is not always the fastest or most efficient way to interact – at times, forcing players to interact through touch can also make their gameplay more difficult. Forcing players to move and interact with their partners in the real world takes time, energy and coordination especially when more than two players are involved. Thus, interpersonal touch can be used to make ordinarily simple tasks more challenging – especially in cases where when players must act quickly.

Though this added challenge may seem like something to avoid, consider the purpose of game design: the goal is not to make a game which is easy, but a game which is fun. Oftentimes it is necessary to challenge players in order to maintain their interest and enjoyment – by challenging your players, you encourage them to strategize, to learn and to grow. As games-theorist Ralph Koster wrote: "true fun is the emotional response to learning" [7].

Interpersonal touch challenges players by forcing them to split their awareness between the digital world and the physical one. In a touch-based game, players must dedicate as much attention to their physical setting and the locations of their partners as the action happening onscreen. Requiring players to coordinate their actions in two separate worlds, creates a new gameplay paradigm – a new cognitive environment to explore and ultimately, a new way to have fun.

iii. Use Touch to Encourage Socialization between Players

Interpersonal touch is an inherently social gesture – at least in the sense that it requires two or more players to occur. An act of interpersonal touch quite literally forms a connection between those involved, and existing research suggests that this connection is as much emotional as it is physical [2, 4, 6]. This socio-emotional connection is of great importance to game design because, for many players, the act of socializing is actually more important than the act of gameplay [8, 9]. Encouraging players to interact through touch, creates a catalyst for ongoing social interaction. A shared touch between two players tacitly breaks down social barriers and encourages further dialogue. For example, in *Prism Squad: GO!* discussions about how and when to blend colors were common among players [10]. Although these conversations were strictly focused on gameplay, we believe that this ongoing dialog also contributed to a social atmosphere between players. Based on our observations, the more players strategized, the more they talked with their teammates, even during the periods of relaxation between stages [10]. When used effectively, interpersonal touch stimulates social interactions between players, and leads naturally towards the "behaviors, rituals, and emotions that make games more exciting." [8]

iv. Manage Player's Social Expectations

Players do not approach your game as a blank slate; for better or for worse, each player will come to you with their own ideas about what interpersonal touch means to them. Making an effective use of interpersonal touch requires designers to anticipate these ideas, and design their games accordingly.

For most players, touch represents connection, attachment, intimacy, or even love. In some situations, this can work to a designer's benefit; a romantically-themed game such as *"Get Lucky" Charms* [3] or *Matchmaker* [11] can use romantic gestures such as handholding as a way to reinforce its message of love and compassion. However, in other game designs, interpersonal touch may be seen as undesirable or even inappropriate – players may feel uneasy about playing games based on touch if they are paired with members of the same sex or players with whom they are not well-acquainted. Your players' cultural and religious backgrounds will also play a role in determining how they feel about interpersonal touch; acts of touch which are permissible in one culture may be stigmatized in another.

We do not believe that this means that interpersonal touch should be strictly relegated to romantically-themed games. But it does mean that game designers must be mindful about their potential audience, and their attitude towards touch. If your players might be uncomfortable with touching their partners, then it may be worthwhile to encourage less intimate forms of interpersonal touch, such as touching on the shoulder or back.

v. Design for the Physical Limitations of Touch

As a physical interaction technique, interpersonal touch is subject to physical limitations which constrain the types of interactions that it can realistically support. The most obvious limitation on interpersonal touch is simply that, in order to for two or more players to touch they must be in close physical proximity. That is to say: a game based on interpersonal touch can only be played by two or more players in a collocated group. However, designers must also be careful not to involve too many players, especially if they rely on a single, shared screen; as more and more players become involved, it becomes more and more difficult for two players to touch at any given time simply because of the increasing number of bodies in the way. When creating games that involve more than two players, designers must be careful to avoid creating situations where one player can interrupt another by reaching for a third. This is especially true for tabletop games, where players may need to reach across the table to touch someone on the opposite side.

Another limitation of interpersonal touch is that it requires at least one hand free. This restriction makes it quite difficult to use interpersonal touch in conjunction with two-handed controllers such as gamepads. This limitation encourages designers to work with "hands-free" interfaces such as tabletop surfaces, or one-handed pointing devices, like the Nintendo Wiimote.

5 Evaluating the Interpersonal Touch Game Design Heuristics

Having outlined our five design heuristics, let us now see how they apply to the four interpersonal touch-based video games which we have previously introduced: *"Get Lucky" Charms*, *Freqtric Game*, *Matchmaker* and *Prism Squad: GO!* By analyzing these games in terms of the heuristics we have presented, we seek to show how our heuristics reflect in practice on the design of interpersonal touch-based games. In the following section, whenever we reflect on a heuristic we will follow it with its heuristic number in parentheses, e.g. (*h-ii*).

5.1 "Get Lucky" Charms

In many ways *"Get Lucky" Charms* [3] makes very effective use of interpersonal touch. It presents a very powerful narrative about touch's role in love, intimacy and sexuality – everything from the title of the project, to the design of the intimate controllers themselves suggests a playful atmosphere where interpersonal touch is very appropriate (*h-iv*). The game is clearly designed for couples and it takes advantage of this design choice by using peripherals which force the players to stand with their bodies pressing against each other (*h-v*). The sensors on the *Intimate Controllers* are positioned to encourage players to touch their partners' breasts and buttocks, which is designed to evoke feelings of intimacy between the players (*h-iii*).

Although it is clear that great care has been spent on the design of *"Get Lucky" Charms'* interface, we feel that less attention has been spent on the design of the accompanying game. In fact, the game itself has very little substance and seems to exist for no other reason than to get the players touching – the gameplay seems to be nothing more than a series of prompts to touch your partner in a specific location on their body. As a result, the game requires absolutely no collaboration between its players (*h-i*). In *"Get Lucky" Charms* each player acts as a dumb receptor for their partner's touch and so players are never actively encouraged to strategize, or to communicate about what they're doing (*h-iii*). Based on these observations, it seems as though the entire point of *"Get Lucky" Charms* is simply to put players in a situation where they can touch each other in an intimate way. That is not necessarily a bad thing – but if that is the case, why make a video game of it? We argue that there are more romantic ways to get close to your significant other – ways which do not even require you to get dressed in sensor-equipped undergarments first.

5.2 Freqtric Game

If nothing else, *Freqtric Game* shows how versatile interpersonal touch-based gaming can be; *Freqtric Shooting*, *Freqtric Dance* and *Freqtric Robot Battle* all approach touch from very different perspectives. In *Freqtric Shooting*, interpersonal touch is very much a cooperative action – players must touch to activate life-saving "bombs" (*h-i*). Since players' supplies of bombs are limited, this encourages players to communicate about when and how to use them (*h-iii*). The mechanic of reviving a downed player through interpersonal touch is somewhat suspect, though; being quickly tapped 100 times in a row by one's partner is more likely to annoy players than anything else (*h-iv*). *Freqtric Dance* uses interpersonal touch to add a new layer to an old classic. Injecting cooperative interpersonal touch actions into *Dance Dance Revolution*'s solo gameplay creates a new challenge wherein players must not just watch the screen, but their partners as well (*h-ii*). These cooperative actions force players to win or lose as a team – players must work together in order to be successful. Finally, the aggressive use of interpersonal touch in *Freqtric Robot Battle* is an interesting way of acknowledging of the role of interpersonal touch in physical competition. Though slapping games are not for everyone, we believe that many male players would feel more comfortable slapping a male partner than they would holding their hand (*h-v*). Such aggressive physical competition may even provide the basis for rewarding socialization – we are reminded of the schoolyard hand-slapping game known as "Red Hands", which, despite its painful consequences, is played in good fun and is often cause for laughter among its players (*h-iii*).

5.3 Matchmaker

Like *"Get Lucky" Charms*, *Matchmaker* [11] is a romantically-themed game. But whereas *"Get Lucky" Charms* puts its emphasis on overt, sexual romance, *Matchmaker* focuses on the "cute" aspects of love. In *Matchmaker*, interpersonal touch comes in the form of handholding – a gesture which is frequently used among couples to demonstrate affection and togetherness (*h-iv*). *Matchmaker* is played directly on the surface of a touch-sensitive tabletop computer, with the players sitting side-by-side – a configuration which easily allows players to hold hands with their partner (*h-v*). Although the game is designed for couples, its relatively inoffensive use of touch means that it could also be enjoyed by other groups of players, such as parents and their children.

In *Matchmaker*, interpersonal touch serves two purposes. The first purpose is to promote an atmosphere of love and romance. In *Matchmaker*, handholding serves as a tangible symbol of the love that the players share. *Matchmaker* encourages players to touch as a way of showing their love for each other (*h-iii*). Interpersonal touch also serves as a form of cooperative interaction between players (*h-i*). Handholding allows players to cure lovelorn Peeps so that they can be matched up again. Curing lovelorn Peeps is crucial part of *Matchmaker*'s gameplay – not only does this mechanic provide much of *Matchmaker*'s challenge (*h-ii*) but it also forces players to communicate and strategize in order to decide when they will hold hands, and when they will match up Peeps instead (*h-iii*).

5.4 Prism Squad: GO!

Although our current implementation of *Prism Squad: GO!* does not recognize interpersonal touch between its players, we are currently working to create an implementation of *Prism Squad: GO!* where players blend colors through touch rather than button-presses. Let us consider how these heuristics would apply to this implementation.

Prism Squad: GO! is a team-based game which places significant emphasis on co-operation between players. In *Prism Squad: GO!*, this cooperation primarily manifests itself through the mechanic of "color-blending". Each player in *Prism Squad: GO!* embodies a particular color, and when two players touch, their colors combine. Interpersonal touch provides players with a simple and direct way to blend colors with their partners (*h-i*). Color-blending (and by extension, interpersonal touch) is very important to *Prism Squad: GO!* – coordinating touch between three players at once provides the game with plenty of challenge (*h-ii*) and serves as a source of ongoing strategic discussion between players (*h-iii*).

Prism Squad: GO! was designed for a broader audience than its predecessor *Matchmaker*, and so concerns over the social appropriateness of touch are much more valid here. Groups of friends who may otherwise enjoy gaming together might be turned off from *Prism Squad* because of its use of interpersonal touch (*h-iv*). In our study of *Prism Squad: GO!*, several participants mentioned that *Prism Squad* would be an enjoyable party game, and we agree that parties are a likely niche for *Prism Squad*. The cooperative aspects of *Prism Squad: GO!* combined with its use of interpersonal touch could make *Prism Squad* an unique icebreaker (*h-iii*).

As a three-player game, *Prism Squad* presents a logistical challenge to working with interpersonal touch. It is very difficult to position three players in front of a screen in such a way that they can all see the screen, but also and can all touch each other at a moment's notice (*h-v*). This logistical challenge could either be a frustrating obstacle, or a fun-but-frantic aspect of gameplay. We believe that the challenge of positing and repositioning your partners so that the appropriate people can touch at the appropriate times could provide a fun challenge in and of itself (*h-ii*).

6 Future Interpersonal Touch Interaction in Video Games

In a previous section, we explored the current state of interpersonal touch-based video games. We have also presented a set of heuristics designed to guide the development of future interpersonal touch-based games. But what does the future of these games look like? How will touch-based interaction in games continue to evolve?

One possibility for future growth involves examining the role of touch-based games in the public social settings. Playtesters have often described games like *Matchmaker* and *Prism Squad: GO!* as "ice-breakers" – fun ways for new acquaintances to get to know each other. With this in mind, we believe it would be very interesting to see how players interact with interpersonal touch-based games in public social gatherings: parties, movie theatres, bars, or even speed-dating venues. Studying how players respond to the presence of these games in a public setting could produce very useful data about where (and to whom) interpersonal touch-based games are best suited.

Another avenue for future research is to explore the use of interpersonal touch in pervasive gaming. Interpersonal touch is a popular physical interaction technique – as such, it plays a key role in many physically-active games such as tag, hide-and-seek, and flag football. Traditionally, video games have lacked the mobility of these physically-active games. But with the popularization of mobile computing and wireless networking, a new class of computer-assisted pervasive video games are emerging, which allow players to play even as they live and move in the real world [5]. As the technology which supports pervasive gaming continues to improve, we believe it would be interesting to explore how interpersonal touch could mediate player-to-player interaction in these pervasive games.

Finally, we believe there are opportunities for further growth in exploring the way in which touch is detected and processed. In existing touch-based games, interpersonal touch is treated as a binary phenomenon: either two players are touching, or they are not. Of course, this is a gross simplification of interpersonal touch: touch can be soft, or it can be forceful, it can be fleeting, or it can linger. Where and how two people touch can also change its meaning; slapping someone on the back is dramatically different from gently patting it. Due to the current state of interpersonal touch-sensing technology, many of these nuances are lost to the game designer. A game like *Matchmaker* cannot distinguish a hug from a handshake [10]. *Intimate Controllers* uses touch-sensors to establish some sense of touch location, but it can only recognize touch on those sensors and nowhere else [2]. A game which could detect not only when, but where and how you touched your partner would have powerful implications for game design; imagine a video game where players must successfully complete a "secret handshake" in order to unleash a coordinated special attack. We believe that such a game would be a lighthearted but amusing way of promoting cooperation and interactivity through interpersonal touch.

7 Conclusion

Interpersonal touch-based games are a recent development in the ongoing evolution of physically-interactive games. Because these games are still so new, details on how to construct and evaluate effective touch-based-games are relatively unknown. In this paper, we have motivated interpersonal-touch based gameplay as a way of promoting inter-player interaction with the goal of supporting increased player socialization. We have also introduced four games which have taken advantage of this interaction technique in their gameplay mechanics: *"Get Lucky" Charms*, *Freqtric Game*, *Matchmaker* and *Prism Squad: GO!*

Based on the lessons we have learned from evaluating these projects, we have produced a set of five design heuristics for the effective use of interpersonal touch in video games. These heuristics are designed to expose the most important considerations for designing games based on interpersonal touch. Using these heuristics as a guide, we examined four existing touch-based games to show how our heuristics can be used as a framework for analyzing interpersonal touch-based games. Finally, we presented our thoughts on the future of interpersonal touch interaction in video games. We hope that this work will prove useful in the development of future games based on interpersonal touch.

References

1. Baba, T., Ushiama, T., Tsuruno, R., Tomimatsu, K.: Video Game that Uses Skin Contact as Controller Input. In: ACM SIGGRAPH 2007 Emerging Technologies (2007)
2. Burgoon, J.K., Walther, J.B., Baesler, E.J.: Interpretations, evaluations, and consequences of interpersonal touch. Human Communication Research 19, 237–263 (1992)

3. Chowdhury, J.: Intimate Controllers (2007), http://www.jennylc.com/intimate_controllers
4. Crusco, A., Wetzel, C.: The Midas Touch: The Effects of Interpersonal Touch on Restaurant Tipping. Personality and Social Psychology Bulletin 10(4), 512–517 (1984)
5. Dreher, T.: Pervasive Games: Interfaces, Strategies and Moves (2008), http://iasl.uni-muenchen.de/links/NAPGe.html
6. Harlow, H.: The Nature of Love. American Psychologist 13, 673–685 (1958)
7. Koster, R.: A Theory of Fun for Game Design. Paraglyph Press, Scottsdale (2005)
8. Lazzaro, N.: Why We Play Games: Four Keys to More Emotion Without Story (2004), http://www.xeodesign.com/xeodesign_whyweplaygames.pdf
9. Voida, A., Greenberg, S.: Wii All Play: The Console Game as a Computational Meeting Place. In: Proceedings of CHI 2009, pp. 1559–1568. ACM, New York (2009)
10. Watts, C., Sharlin, E., Woytiuk, P.: Exploring Interpersonal Touch-Based Interaction and Player Socialization in Prism Squad: GO! In: 3rd Annual International Conference on Computer Games, Multimedia and Allied Technology, Singapore (2010)
11. Watts, C., Sharlin, E., Woytiuk, P.: Matchmaker: Interpersonal Touch in Gaming. In: Chang, M., Kuo, R., Kinshuk, Chen, G.-D., Hirose, M. (eds.) Learning by Playing. Game-based Education System Design and Development. LNCS, vol. 5670, pp. 13–24. Springer, Heidelberg (2009)

Investigating the Affective Quality of Motion in User Interfaces to Improve User Experience

Doyun Park and Ji-Hyun Lee

Graduate School of Culture Technology, KAIST, 335 Gwahak-ro,
Yuseong-gu, Daejeon 305-701, Republic of Korea
{soulvibe,jihyunlee}@kaist.ac.kr

Abstract. This study focuses on *motion* in user interfaces as a design element which can contribute to an improved user experience of digital media entertainment. The design for user experience is necessary to deal with user emotion especially in the entertainment domain. As a means to approach emotion, we studied affective qualities which are the features of an artifact that can influence emotion. In user interface design, motion has not been practically dealt with in this perspective. Through empirical study, we verified that motion plays a significant role in forming the affective quality of user interfaces and found that content type and application type has influence on this effect. Moreover, a preliminary investigation was made on the use of the Effort system from Laban's theory for the design of motion in terms of affective quality.

Keywords: Motion, Affective quality, User interface, User experience.

1 Introduction

The emergence of user experience (UX) and diversifying entertainment content through digital media has brought new challenges to user interface design. Previously, the job of user interface designers was to ensure good usability so that users could operate the system at ease. But in the UX paradigm of today, it has become important to design for better experience by providing a holistic experience with consideration to emotion. This has greater significance in the entertainment domain than in most other areas because users fundamentally seek entertainment for an experience.

In order to approach emotion, designers need to consider the affective outcome of their design. According to Russell (2003) the features of an artifact that influence user's emotion is known as *affective quality* [1]. This refers to the feeling and impression that an artifact has which can cause changes in the user's emotion, such as the distinguishable feeling that Apple products convey. Therefore, for a better UX, the affective quality of every element in an artifact needs to be considered in design.

In this perspective, the focus of this study is on the design of motion in user interfaces. Motion has become an essential element of user interfaces through enhanced processor speed and bandwidth. Most interactive media, such as websites, IPTV, portable media players and mobile phones nowadays display many motions. However, unlike the static elements such as color and layout, the affective quality of motion has not received much attention. This is because motion was mainly used for instrumental purposes, such as giving feedback or attracting attention [2], [3].

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 67–78, 2010.

With the growing use of motion, its relationship with the affective quality of user interfaces needs to be studied as a way to improve UX. Hence, the aim of this study is to first verify the influence which motion has on the affective quality of user interfaces through empirical investigation, and secondly investigate a method to design motion in user interfaces by applying a relevant framework.

2 Related Work

There have been many studies which dealt with the relationship between movement and emotion. Vaughan (1997) described the emotional effects that movement can have and the need to understand movement to enhance user's experience [4]. Bacigalupi (2000) also emphasized the need to handle movement with caution and suggested a conceptual framework for the design of movement in multimedia [5]. These studies provide the grounds for understanding movement, but the practical methods to actually design them for emotion are left to be studied in the context of user interfaces.

Studies on the methods to design movement for emotion have been conducted more frequently in the product design field. Lee (2007) studied the relationship between physical movement and emotion for application in interactive products, while Saerbeck (2010) looked into the emotional responses that robots can give with their movements [6], [7]. These studies reveal insights on how physical movement influence emotion. Nevertheless, the scope is often limited by examining the relationship of movement directly to emotion.

To gain a better understanding of how the movement element influences emotion, the relevant affective qualities need to be investigated as an intermediate step. This is because, as mentioned above, affective quality is one of the key factors which causes change in the user's emotion. Thus, to design elements in user interfaces towards an intended emotion, we need a practical and systematic method to control the elements in relation to affective quality.

3 Theoretical Background

3.1 Framework of Movement

The conceptual framework on movement by Bacigalupi (2000) describes two main qualities of movement [5]. One is the formal quality which includes rhythm, tempo, sequence, and direction. The other is the expressive quality which is composed of kinesis and empathy. The formal quality determines the form that the movements take while the expressive quality determines how each movement is expressed.

For our investigation on affective quality, we will focus on the expressive qualities of movement. The two elements of the expressive quality, kinesis and empathy, provide the frameworks which we can begin our study. Kinesis is about how a movement is expressed and it is described by the Effort system of Laban's theory. Empathy, on the other hand, is about the emotional engagement or physical identification related to a movement.

The relationship between kinesis and empathy provides a way to design for the affective quality of movement. By controlling kinesis, we can realize empathy accordingly. In this respect, the components of kinesis are reviewed in the next section for our study.

3.2 Laban's Theory of Effort

Laban Movement Analysis (LMA) is a method developed by Rudolf Laban to interpret, describe, visualize and notate human movement. It is generally used to analyze the movement of dancers and athletes.

Laban's theories comprise of four main categories: Body, Effort, Shape, and Space[1]. Out of these, Effort is a system used to understand the characteristics of the way a movement is performed with respect to inner intention. Thus it is closely related to the description of movement in terms of its quality and expressiveness. The Effort system is divided into four factors, each of which has two opposing elements as shown in table 1 [8].

Table 1. Description of the Effort system

Effort Factor	Element	Description
Space	Indirect	flexible, meandering, wandering, multi-focus
	Direct	single focus, channeled, undeviating
Weight	Light	buoyant, easily overcoming gravity, decreasing pressure
	Strong	powerful, having an impact, increasing pressure
Time	Sustained	lingering, leisurely, indulging in time
	Sudden	hurried, urgent
Flow	Free	uncontrolled, abandoned
	Bound	controlled, restrained

Most movements show a combination of these factors. A combination which often occurs is of Space, Time and Weight. For example, a combination of Direct, Strong and Sudden elements would be embodied in a punching movement, while Indirect, Light and Sustained elements will result a floating movement. The three factors can combine to express eight different types of movement (Table 2) [9].

Table 2. Eight combinations of Space, Time and Weight factors

Movement type	Space	Time	Weight
Slashing	Indirect	Sudden	Strong
Gliding	Direct	Sustained	Light
Pressing	Direct	Sustained	Strong
Flicking	Indirect	Sudden	Light
Wringing	Indirect	Sustained	Strong
Dabbing	Direct	Sudden	Light
Punching	Direct	Sudden	Strong
Floating	Indirect	Sustained	Light

[1] The terms referring to the theories of LMA are capitalized to distinguish them from their commonly used meanings.

Although the Effort system was devised for the study of human body movement, it provides a systematic way to describe the expressive quality of movement with the four factors. Thus, it is an effective and practical framework to work with. In our study, we will investigate the use of the Effort system for the design of motion in user interfaces.

4 User Study on Motion and Affective Quality

A user study was performed to verify the significance of the motion element on the affective quality of user interfaces. This user study aims to investigate the following points:

- The influence of motion on the affective quality of user interfaces.
- The dependency of this influence on two contextual factors: content type (e.g. the genre of the content) and application type (e.g. interface for music, movies, photos, etc.).

4.1 Prototypes

Three different interfaces were developed for the user study by Adobe Flash CS3: a minimal interface showing a moving object (Fig. 1) and two image viewing interfaces with different contents – one with soccer match images and the other with tropical resort images (Fig. 2). The minimal interface and the image viewing interfaces were developed to investigate the influence of application type. As for the image viewing interfaces, two distinctly different contents were implemented to investigate the influence of content type.

Three types of motion were applied to these prototypes. Two of these motions were designed to display opposing combinations of the Space, Time, and Weight factors of the Effort system. With the parameters adopted from Chi (1999), motion A had parameter settings for Direct, Sudden, and Strong elements while motion B had

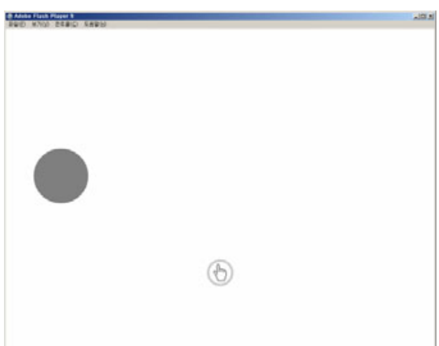


Fig. 1. The minimal interface displays an object that shows motion at user input. It consists of minimal elements to make the users focus on motion. It. Different types of motion are shown as the object changes position on the x-axis.

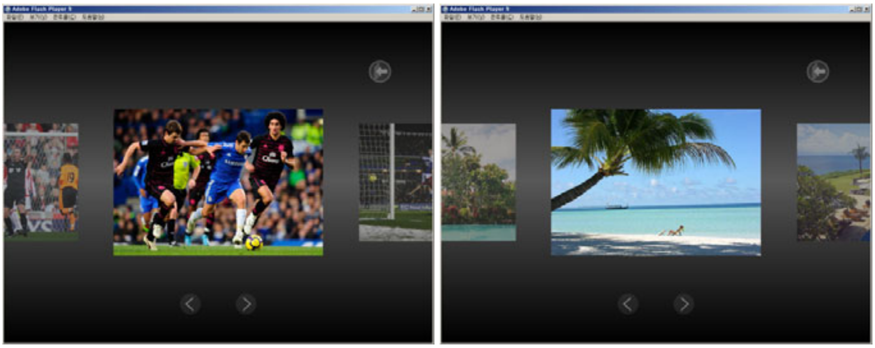


Fig. 2. The image viewing interface allows users to browse through a set of photos as they shift horizontally from one to another on input. Two types of content were implemented in this interface: one that shows soccer match images (left) and the other that shows tropical resort images (right).

Table 3. Motion parameters applied to prototypes

Property	Motion A	Motion B	Motion C
Path curvature	Straight	Curved	n/a
Duration	0.4 sec	3.6 sec	0 sec
Acceleration	Acceleration	Deceleration	n/a
Anti./Oversh.	Anticipation	Overshoot	n/a

parameter settings for Indirect, Sustained, and Light elements (Table 3) [8]. Only the four properties which were generally applicable on an abstract object were applied: the path curvature property determines the straightness or roundness of the path that the object moves through, the duration property determines the length of time the motion occurs, the acceleration property determines the rate of change in velocity, and the anticipation/overshoot property determines if there is a slight pull before the motion begins or if there is a slight swing out at the end of the motion.

Motion C is a common type where an object changes position instantly without any motion in-between. This motion was implemented to see how much more influence dynamic motions can have over the common minimal use of motion. All prototype interfaces were presented with Adobe Flash Player on a 20 inch touch screen monitor. User input was made by tapping the objects on the screen.

4.2 Experiment Design

In our experiment, motion was the first independent factor and context was the second independent factor. To investigate the influence of these two factors, the perceived affective qualities were measured for the three types of user interfaces, each of which displayed the three types of motions mentioned above. A within-subject design was used to reduce the possible influence of individual differences.

For measurement, the following seven types of affective qualities were surveyed on a five point Likert scale: refreshing, calm, luxurious, powerful, youthful, uniqueness and futuristic. These affective qualities were adopted from Hong (2009) where the seven types were classified as the prevailing affective qualities in website interfaces [10].

4.3 Procedure

20 graduate students, 12 male and 8 female, aged from 23 to 29, participated in the experiment. First, the participants rated the affective quality of the three motion types displayed by the minimal interface. Next, they experienced and rated the image viewing interfaces: three sessions for each motion types on both the soccer match images and tropical resort images.

For the image viewing interface, different image sets were presented in every session to prevent bias from seeing same images repeatedly. At the same time, to cancel out the influence of the differing images on the evaluation of affective quality, the images within each content type were selected from identical sources and the image sets were paired randomly with motion types for every participant.

4.4 Result

The ratings on the affective qualities varied according to motion type in all three interfaces. Multivariate analysis of variance (MANOVA) was used to analyze the influence of the two independent factors on the set of affective qualities. We could verify that the motion factor had a significant influence on all tested affective qualities in all three interfaces at the $p<.05$ level. On the other hand, the contextual factors (content type and application type) only showed significant influence on some affective qualities for some motions.

In the soccer match viewing interface, motion B significantly lowered the *refreshing, powerful*, and *youthful* affective quality than the other motions while significantly increasing the *calm* affective quality of the interface (Fig. 3). On the other hand,

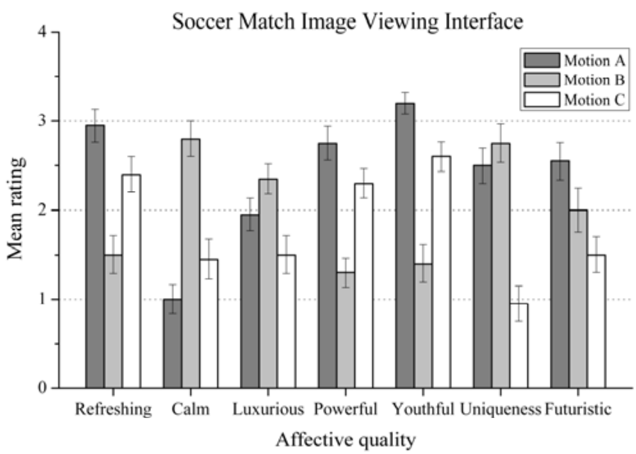


Fig. 3. Mean ratings on the affective qualities of the soccer match image viewing interface show statistically significant differences according to motion type. Error bars represent the standard error of the mean.

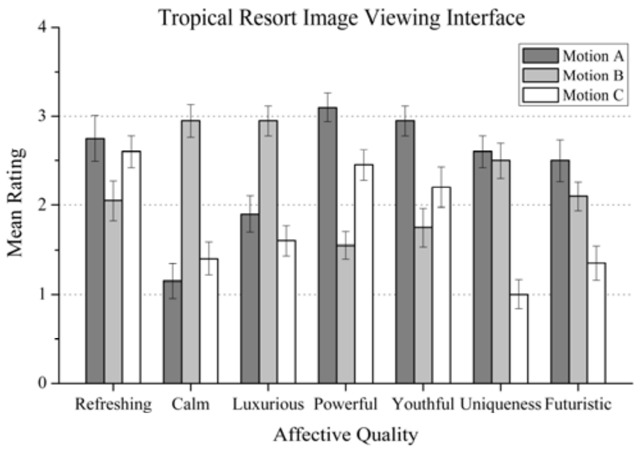


Fig. 4. Mean ratings on the affective qualities of the tropical resort image viewing interface show statistically significant differences according to motion type.

motion A significantly increased the *youthful* affective quality while motion C lowered the *uniqueness* affective quality in comparison to other motions. Similar influence was found from the tropical resort content (Fig. 4).

Content type showed little significance on how motion influences the affective quality of user interfaces. The ratings of each motion were compared among the soccer match and tropical resort image viewing interfaces. Only motion B showed a significantly differing level of influence on the *luxurious* affective quality by strengthening it in more the tropical resort content than in the soccer match content. Nevertheless, the rank order of motion type on the level of influence for each affective quality was generally identical among the two content types.

Next, the affective qualities measured from the minimal interface were examined. The three motions used in this experiment showed statistically significant influence on all affective qualities at the $p<.05$ level except *uniqueness*. In particular, motion B in this interface significantly lowered the *refreshing, powerful* and *youthful* affective qualities and significantly increased the *calm* and *luxurious* affective qualities than other motions (Fig. 5).

As a result, application type had greater significance on the influence of motion than content type. Rating comparison between the image viewing interfaces and the minimal interface showed significant difference in how much motion influenced the *calm, luxurious, powerful, uniqueness,* and *futuristic* affective qualities. Moreover, the rank order of motion type on the level of influence was not identical to the image viewing interfaces for the *uniqueness* and *futuristic* affective qualities.

4.5 Discussion

From the above results and statistical analyses, we were able to verify that the motion element has a significant influence on the seven affective qualities of user interface that were tested.

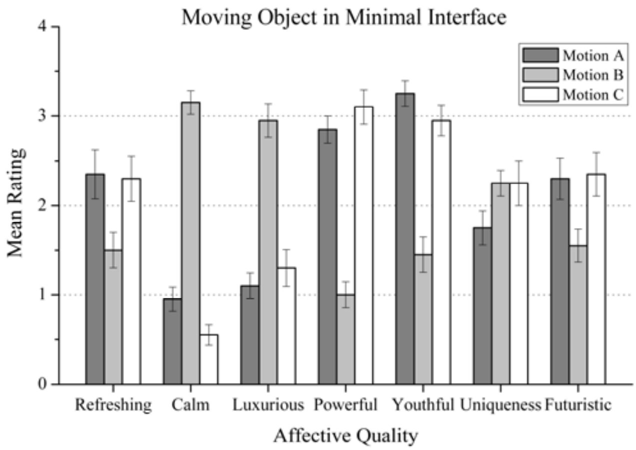


Fig. 5. Mean ratings on the affective qualities of the moving object show statistically significant differences by motion type except for the uniqueness affective quality.

Content type showed partial influence on how motion affects affective quality. Motion B had greater influence on the *luxurious* affective quality in the tropical resort content than in the soccer match content. This difference is analyzed to be the result of the users' attitude being more accepting to motion B within the tropical resort content than in the soccer match content because its leisurely feel is expected from the relaxing resort images. From this analysis, we can understand how content type acts as a variable. However, this influence is not so profound as to change one motion to have a dominant effect in one content type and not in another.

Lastly, the influence of application type was examined by comparing the results from the minimal interface and the image viewing interfaces. The overall result from the moving objects was similar to the image viewing interfaces. However, the influence of motion C on the *uniqueness* and *futuristic* affective quality was significantly greater in the moving object interface. This difference is analyzed to result from the user's differing expectations on each application types. The *uniqueness* and *futuristic* feel by motion C was not so significant in image viewing interfaces because users were accustomed to photo images shifting instantly in websites.

Our analysis shows that motion should be considered as a factor of affective quality in user interfaces. Moreover, the content type and application type requires examination in the user's perspective to accurately predict the effect of motion.

5 User Study on the Properties Related to the Effort System

The verified significance of motion provides motivation for a practical method to design motion. To do this, we began an investigation on the use of the Effort system of Laban's theory because, by being able to control motion in terms of this framework, the affective qualities of motion can be controlled and communicated more effectively. As a preliminary stage of this investigation, we performed a user study to identify the properties in a user interface that relate to three factors of the Effort system in comparison to human body movement.

5.1 Prototype

A minimal interface that displays eight different types of motion through a moving object was developed using Adobe Flash CS3. Each motion was designed to be a combination of the Space, Time and Weight factors of the Effort system (punching, floating, gliding, wringing, pressing, dabbing, slashing and flicking).

The parameters for each combination were adopted from the observations made on human body movements in [8]. The parameters applied were related to four motion properties which are generally applicable on an abstract object in user interfaces: path curvature, duration, acceleration, and anticipation/overshoot (Table 4). The applied level of the acceleration and anticipation properties varied for some Effort combinations according to the observations in [8]. All motions were presented with Adobe Flash Player on a 20 inch touch screen monitor. The object displayed motion when the participant tapped the object.

Table 4. Motion parameters applied for each element of the three Effort factors

Property	Space		Time		Weight	
	Indirect	Direct	Sustained	Sudden	Light	Strong
Path Curvature	curved	straight	-	-	-	-
Duration	-	-	long	short	-	-
Acceleration	-	-	decelerate	accelerate	decelerate	accelerate
Anti./Oversh.	-	-	overshoot	none	overshoot	anticipation

5.2 Experiment Design

We tested four motion properties which were the independent factors and measured how the three Effort factors were perceived. For measurement, five point semantic differential scales were used on each Effort factors: the Space factor ranging from Indirect to Direct, the Time factor ranging from Sustained to Sudden, and the Weight factor ranging from Light to Strong.

5.3 Procedure

21 graduate students, 10 male and 11 female, aged from 23 to 32, participated in the experiment. Eight motions representing the Effort combinations were shown to participants in random order. After sufficiently viewing the motion, participants rated how it was perceived in terms of the three Effort factors.

5.4 Result

For all the motions tested, the Space and Time factors were perceived as expected by the parameters applied. With the Weight factor however, the parameters applied for *dabbing*, *flicking*, *pressing* and *wringing* motions were perceived opposite to what was expected. The parameters applied for *dabbing* and *flicking* were perceived as Strong while the parameters applied for *pressing* and *wringing* were perceived as Light.

MANOVA was performed to identify the effect of the four properties that were applied. Path curvature had a significant influence on the Space factor while duration and acceleration had significant influences on the Time and Weight factors at the $p<.05$ level. On the other hand, anticipation/overshoot did not have significant influence on any of the three Effort factors (Fig. 6).

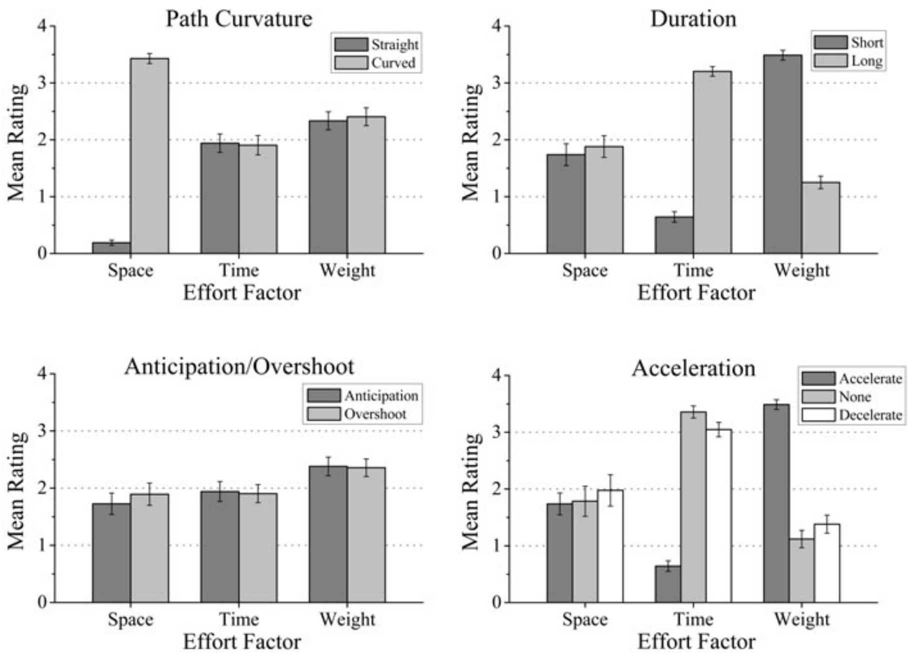


Fig. 6. Mean ratings of each Effort factors relative to the parameters of the four properties. Ratings close to 0 were perceived to be Direct, Sudden and Light, while values close to 4 were perceived to be Indirect, Sustained and Strong for each of the Space, Time and Weight factors, respectively. The error bars represent the standard error of the mean.

5.5 Discussion

From this user study, we were able to identify the motion properties that can be used to control motion in terms of the three Effort factors in user interfaces. In coherence with human body movement, path curvature was related to the Space factor, duration was related to the Time factor, and acceleration was related to both Time and Weight factor in user interfaces.

Differences were also discovered. In human body movements, the anticipation/overshoot property was observed to be related to the Time and Weight factor. However, it did not show relation to any factors in the user interface setting. On the other hand, the duration property showed strong relation to the Weight factor in user interfaces while this relation was not observed in human body movement. This explains why four of the motions types tested were perceived as having the opposite element on the Weight factor.

Unlike human body movements, we can assume that duration and acceleration plays a significant role in the perception of the Weight factor in user interfaces. These findings imply that the properties of motion act differently on the perception of affective quality in user interfaces when compared to human body movements.

6 Conclusion and Future Work

Through empirical study, we confirmed that motion significantly contributes to the affective quality of user interfaces. The level of influence by motion varied according to the content type and application type for several affective qualities. This effect is likely to have come from varying user expectations according to the two contextual factors. Thus, user expectation must be studied when designing motions for the affective quality of user interfaces. Overall, our result provides a motivation towards the study of motion design for affective quality.

Our second study identified the motion properties related to the three Effort factors in user interfaces and compared it with the relations found in human body movements. The investigated properties showed several differences in how they influence the Effort factors in user interfaces. These results imply that it is inadequate to directly adopt the human body movement parameters of the Effort system for motions in user interfaces. Further investigation on the properties is required to feasibly apply the Effort system for the design of motion in user interfaces.

For future work, an in-depth study on how the motion properties relate to the Effort factors is required. Moreover, studies on a wider range of users are required for practicality, since our experiments were performed on a limited number and cultural range of participants. Also, an investigation on the formal qualities of motion would be necessary to expand the variety of affective qualities designable in user interfaces. This knowledge will provide leverage in designing motions for satisfying affective qualities in user interfaces which can improve the UX of entertainment in digital media.

References

1. Russell, J.A.: Core Affect and the Psychological Construction of Emotion. Psychological Review 110, 145–172 (2003)
2. Stone, D., Jarrett, C., Woodroffe, M., Minocha, S.: User Interface Design and Evaluation. Morgan Kaufmann, Amsterdam (2005)
3. Galitz, W.: The Essential Guide to User Interface Design. Wiley, New York (2002)
4. Vaughan, L.C.: Understanding Movement. In: SIGCHI conference on Human factors in computing systems 1997, pp. 548–549 (1997)
5. Bacigalupi, M.: Designing Movement in Interactive Multimedia. Interfaces 44, 12–15 (2000)
6. Lee, J.-H., Park, J.-Y., Nam, T.-J.: Emotional Interaction through Physical Movement. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4552, pp. 401–410. Springer, Heidelberg (2007)

7. Saerbeck, M., Bartneck, C.: Perception of Affect Elicited by Robot Motion. In: 5th ACM/IEEE international conference on Human-robot interaction, pp. 53–60 (2010)
8. Chi, D.M.: A Motion Control Scheme for Animating Expressive Arm Movements. Ph.D. Dissertation, University of Pennsylvania (1999)
9. Laban, R., Lawrence, F.C.: Effort: Economy in Body Movement. Plays, Inc., Boston (1974)
10. Hong, S.-Y., Lee, H.-J., Jin, K.-N.: A Study on the Image Scale through the Classification of Emotion in Web Site. Science of Emotion & Sensibility 12, 1–10 (2009)

The MINWii Project: Renarcissization of Patients Suffering from Alzheimer's Disease Through Video Game-Based Music Therapy

Samuel Benveniste[1], Pierre Jouvelot[1], and Renaud Péquignot[2]

[1] CRI MINES ParisTech, 35, rue Saint-Honoré, 77305 Fontainebleau, France
Tel/Fax: +33 1 64 69 47 08
{samuel.benveniste,pierre.jouvelot}@mines-paristech.fr
[2] Hôpital Saint-Maurice, 14, rue Val d'Osne, 94415 Saint-Maurice Cedex, France
Tel/Fax: +33 1 43 96 63 63
r.pequignot@hopital-saint-maurice.fr

Abstract. MINWii, a new serious video game targeting Alzheimer and demented patients, is a simple Music Therapy tool usable by untrained care givers. Its objective is to improve patients' self-image (*renarcissization*) to reduce behavioral symptoms, which are an important cause of institutionalization. With MINWii, elderly gamers use Wiimotes to improvise or play predefined songs on a virtual keyboard. We detail our design process, which addresses the specific features of dementia: this iterative refinement scheme, built upon qualitative, small scale experiments in a therapeutic environment, led to a shift of MINWii's original focus from creativity to reminiscence. A large majority of our patients, with mild to moderate dementia, expressed a strong interest in our system, which was confirmed by feedback from the care givers. A fully controlled usability study of MINWii is currently under way, which should lead to future research assessing its actual therapeutic impact.

Keywords: Alzheimer, Dementia, Music Therapy, Wiimote, Renarcissization.

1 Introduction

According to the World Health Organization [20], more than 18 million people worldwide suffer from Alzheimer's disease, the most common form of dementia. Researchers estimate that this pandemic may easily escalate to affect 1 in 85 people in 2050 [6], 70% of whom in developed countries. Caring for so many dependant patients already costs trillions of dollars annually [20], and remains rather uneffective. Indeed, the best that current drugs can do, in most cases, is to slightly slow down the cognitive decline.

Consequently, until satisfying treatment methods are found, our aging societies will have to adapt and attend to the needs of an ever-growing number of demented patients; non-pharmaceutical approaches can help with this issue [2].

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 79–90, 2010.

Among them, two techniques seem to stand out [13]: Cognitive Stimulation Therapy (CST) [18] and the one of interest here, namely Music Therapy (MT).

MT techniques fall into two broad categories: Receptive MT, where patients listen to musical pieces under the supervision of therapists, and Active MT (AMT), where patients actually compose and perform music. Since patients treated with MT often lack any musical training and cannot properly play anything beyond simple percussions, elaborate computer systems such as Hyperscore from MIT Media Lab[5] or Riley's musical touchscreen [16] have been designed to address this issue, although their use has yet to spread in the MT community.

Here, we experiment with a new approach: given the surprising success of Nintendo's Wii with the elderly [21], we borrow techniques and off-the-shelf components such as the Wiimote from the world of video games to empower MT practitioners with a low-cost yet versatile tool, as was done in previous work [4], albeit for a very different condition (ADHD). To create MINWii, where patients use Wiimotes to either improvise music or play predefined songs on a virtual keyboard displayed on screen, we took the specific features of Alzheimer's disease into account when designing the gameplay. We then performed a qualitative, small scale usability test in a therapeutic environment to iteratively improve our design. We show here that the system is indeed usable by care givers with no musical training dealing with patients with mild to moderate impairment; moreover, a large majority of our test patients expressed a strong interest in MINWii, which was further confirmed by feedback from the care givers.

In Section 2, we explain our renarcissization-based approach through the use of video games and MT. We motivate the key design choices we made in MINWii in Section 3: low cognitive and motor HMI requirements, failure-free gameplay and use of off-the-shelf components. We describe in Section 4 our two-phase experimental validation process, where we test our first design, implement subsequent upgrades and show how the final version helps patients improve their self-image. We conclude and introduce future work in Section 5.

2 Renarcissization for Dementia

When used in the context of dementia treatment, renarcissization [1] refers to the process of restoring self-esteem in patients who have gradually grown to consider themselves as a useless burden for their care givers. Our approach gives them the opportunity to engage into activities specifically designed to highlight and make use of whatever capabilities they may still have.

2.1 Behavioral Issues

Besides reducing patients' depression levels, renarcissization could have a significant economic impact since, in an overwhelming majority of cases, behavioral issues are one of the main causes of institutionalization, often against the patients' will. Indeed, they frequently try to prove to their care givers that they can still live their life as usual, failing to realize the extent of their disability. They might for example routinely take their shower with their clothes on or completely mess up their home while trying to clean, exhausting their care givers. Worse, they might jeopardize their own safety and that of others by imprudence, eventually leading their family to give up and send them to a hospital or an institution, where they are more likely to be depressed and die faster [19].

On the contrary, patients with higher self-esteem are less likely to try to do things they cannot do anymore. A large part of them, who are otherwise healthy enough to stay at home, therefore require only minimal daily care such as food delivery and help for personal hygiene. This has great benefits in terms of cost: a visit from a nurse and a few hours of help from a personal assistant every day cost roughly half as much as a permanent stay in a caring home and almost three times less than a stay at the hospital.

Thus, there is a dire need for affordable and easy-to-use tools that untrained helpers such as family members could use to reduce or even prevent the onset of behavioral disorders in demented patients with minimal equipment and training costs. We expect that the blending of MT and video games within MINWii can (1) be instrumental in making demented patients feel better and (2) ease the time-consuming, psychologically-demanding burden of caring for them on their assistants and families.

2.2 Music Therapy and Video Games

The rationale for using MT in the treatment of Alzheimer's disease is as simple as it is compelling: for a large majority of patients, musical sensitivity, in particular regarding songs discovered during their childhood, is one of the very last cognitive capabilities to disappear. Moreover, it has been shown that techniques such as U-Cycle Receptive MT [9] efficiently reduce the anxiety levels of mild to moderately incapacitated patients. Thus, given its power and universality, music appears to be a good potential vector for the renarcissization of demented patients.

However, incorporating AMT in the treatment of dementia can be difficult: (1) there is a lack of music therapists even though demand is growing thanks to influential supporters such as Oliver Sacks [17] and (2) patients usually cannot use elaborate instruments. As a result, traditional AMT sessions, though in fact therapeutically effective, often end up as somewhat unappealing to patients and their family, with the therapist doing most if not all of the playing.

We think using a video game-based approach can alleviate at least some of these problems:

- By building on the success of the Wii in elderly care institutions [21] and providing data logging for easy assessment of patients' progress, we intend to convince geriatricians to consider and support the use of AMT;
- By providing ways for non-MT practitioners to use AMT in virtually any setting, including the patient's home, since MINWii simply requires a computer and a TV, we hope to help establishing AMT as a standard practice in dementia treatment and, hopefully, encourage institutions to go one step further and hire a real music therapist;

– By offering a virtually infinite library of customizable instruments and interfaces, we aim at helping AMT become more enticing and patient-specific.

3 Designing for Dementia

Demented patients' skills deteriorate at surprisingly different rates. Some simple aptitudes such as the ability to remember a few objects may be long gone before seemingly higher-level ones such as singing a song begin to decline. Therefore, a straightforward simplification of a regular music video game like Guitar Hero [10] would probably not be an effective music therapy tool. We suggest below that the playing experience can be made much more enjoyable for the patients if, instead, we manage to build upon the specific abilities they still have while keeping very low requirements in the areas where they are diminished. In this section we describe the key design points we established for the first version of our system: low cognitive and motor requirements, a rewarding user experience and hardware, software and operational simplicity.

3.1 Low Cognitive and Motor Requirements

Conditions such as arthritis and motor coordination deficits prevent elderly patients from making large or fast movements involving too many joints. Therefore our design emphasizes pointing using Wiimotes: only wrist movements within most patients' zone of comfort in terms of amplitude are required. Furthermore, MINWii heavily filters Wiimote inputs and cursor movements so that the pointer behaves smoothly enough for the patients to feel comfortable. Finally, to make it easier to hold the Wiimote, we added the possibility to use a Wii Pistol: this hollow plastic gun, in which the Wiimote can be encased, enables playing without having to bend the wrist in a way that might be uncomfortable. Regarding cognitive issues, pointing has a low footprint, since it is very intuitive: the user does not need to learn a new convention such as "this button does this" or "this arrow does that". Moreover, the Wii Pistol obviously has a great affordance for this task, though it has the potential downside of diverting the focus of the game from music to shooting.

To keep lowering the cognitive weight of our system, we designed an extremely simple, image-based graphical interface. We were very careful to adapt the visual part of the application to our population's reduced attention and focus abilities, often complemented with some degree of visual impairment. Everything on the screen is very large and highly contrasted to be easy to comprehend and we kept the eye candy to a minimum to avoid distracting users with unnecessary frills.

More to the point, we also tried to accommodate for the great decision-making difficulties demented patients usually have. The buttons on the Wiimote all do exactly the same thing and there is always one and only one active area at a time, highlighted as clearly as possible. This way, with only one method of interaction and only one object to interact with, the user never has a choice to make once the game has started and is much less at risk of feeling lost. Interestingly, this

Fig. 1. Instrument choice screen

feature is at the opposite of what is usually considered as a key gameplay asset in traditional video games; the challenge here is to shrink the choice space while keeping the fun factor as high as possible.

Finally, to make the system look easy to use, we chose light pastel colors to get a nice soothing look, since it has been shown that a beautiful interface actually is easier to use [11]. Also, all the configuration settings (game difficulty, Wiimote sensitivity, pointing mode,...) are made by the care giver with his/her own Wiimote with no graphical output; indeed, we observed that visual indications related to such abstract notions often confuse demented patients and might induce anxiety by making the system look too complicated (see below).

3.2 Failure-Free Gameplay

Demented patients often exhibit a systematically defensive attitude regarding anything unfamiliar. They are always afraid to fail or break something when confronted with a new task. They will also give up very easily if they do not succeed right away, even when receiving strong encouragements from the care givers.

For such reasons, we designed a gameplay where failure is either impossible or very unlikely due to the simplicity of the task at hand. Moreover, if it does happen, its importance is systematically toned down. First, no click will result in a dissonant sound since only notes guaranteed to sound well together can be played during the game (e.g., a C major scale). Second, given how difficult it is for demented patients to use almost any electronic device autonomously whatever the design efforts invested [16], we prefer to see our system as a tool to foster interaction between patients and care givers. MINWii does not present judgments of any kind (scoring, game over, etc.) to the patient and relies on the care giver to both give praise and help through failure.

3.3 Simplicity in Design

As stated before, we have chosen to rely exclusively on off-the-shelf components in order to build a system that could be installed pretty much anywhere for less than $1000, which is roughly the average price of a one-day hospital stay. The only necessary equipment is: a standard computer ($500), a computer-enabled TV screen ($400), a wireless infra-red Sensor Bar ($20) to be placed below the screen and at least one Wiimote ($40). We did use commercial sound synthesis PC software for prototyping, but the next version of MINWii, expected to be soon used in a large scale test, will rely exclusively on free software (pygame [15], fluidsynth [7] etc.) under the GNU Public License. Note that all the interactions can be logged for subsequent analysis. Though it is not required, we suggest using a personal USB thumb drive ($15) for each patient, containing his/her prefered configuration, logs and scores throughout the treatment, as is done at InGame Lab [3].

Off-the-shelf components like the Wiimote have another interesting characteristic, less obvious but probably even more relevant: they usually have a simple, pleasant aesthetic design. For demented patients who are typically over 80 years old, this is of importance since any apparent complexity such as lots of wires, strange-looking machines etc. may trigger their constant fear of failing and breaking objects. The experience of one of us with personal hygiene robots [14] showed that even carefully designed, very easy to use devices can be of little practical interest with some patients because of that. MINWii only uses two pieces of nowadays familiar equipment (laptop and TV) and remote-like devices (Wiimotes and Sensor Bar); it is thus much less likely to be rejected for this kind of reasons.

3.4 MINWii 0.9

The initial version of MINWii we designed included only the Improvisation and Challenge Modes. Both can be played with any of the nine instruments available; note that we could have included many more and chose to offer only nine just to keep the pictures big enough on the screen (Figure 1).

Improvisation Mode. Here, patients are invited to improvise using a scale of their choice (e.g. C major) by pointing at a virtual keyboard of 8 or 11 colored keys displayed on the screen, depending on their level (determined by the care giver). A large white dot shows where they are aiming at and highly contrasts with the keys in the background, which all have their unique pastel color and display the name of their note (in French). The closer to the top of the screen one clicks, the higher the volume is (see Figure 2). Though it is hard to play fast, rather elaborate melodies can be played quite expressively at a slow tempo. However the temptation to speed up can be great, so we added a *glissando* option that can be switched on or off by the care giver; this allows for more "experimental" improvisations, but makes it much more difficult to play expressively.

Challenge Mode. Here, the patient has to choose a song to play. We started with the songs that patients tend to remember the most, hoping it would encourage them to play: simple nursery rhymes that most French children are familiar

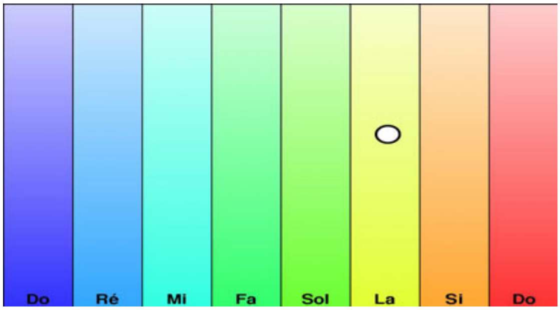


Fig. 2. Improvisation screen

with. This is a general rule with Alzheimer's disease: childhood memories tend to remain longer than more recent ones and therefore will grow in importance as patients forget the rest.

In this mode, the patient clicks the keys that highlight successively according to the song chosen. If the player hits a wrong key, the corresponding note is played, but at a lower volume. Since all the notes belong to the same scale, this will usually not sound too bad and can even allow the player to add a little personal touch to the song if he/she wants.

4 Testing and Results

We used a protocol akin to Action Research [12] to test MINWii and freeze the final design of our system while already helping patients at the same time. By putting these often lonely people at the center of the design process and making them feel useful by asking them for opinions, giving them praise or simply thanking them, we hoped to compensate for engaging them in an activity that we had yet to prove to be beneficial and that could even conceivably have been detrimental to their well-being, however unlikely that seemed to us at the time.

4.1 Population and Protocol

MINWii primarily targets demented patients with mild to moderate cognitive impairment, and potentially more severe (but not profound) difficulties in the visual, auditive and motor areas. We tested our system in the Geriatrics Unit of Hôpital Saint-Maurice (France), where patients usually stay for up to two months, depending on their condition and the follow-up care available.

We selected 9 patients with suspicion of Alzheimer's disease, a Mini Mental State score [8] between 10 and 25 (out of 30) and no profound deafness, blindness or motor disabilities. From November 2009 to February 2010, once a week, available patientswere encouraged to come to the animation room and "play music with the remotes". We usually ended up with a group of 3 or 4 whom we would sit in a semi-circle in front of the projection screen. They would each be given a

Wiimote and then would take turns playing the song of their choice, with staff members sitting next to them and giving as much help as needed.

4.2 A Need for Guidance

For the very first session, only the Improvisation Mode was available. We were thrilled to see that patients were willing to play when asked to, some of them even showing signs of great enthusiasm and rather surprising pointing skills for a first trial. However, none of them were willing to explore the system in depth on their own, let alone improvise. When invited to do so, they said things like "I don't know any music" or "Hum me some tunes" and never even tried to click random notes, shake the Wiimotes or click the corners of the screen the way most of the healthy subjects who tried the application during informal testing did. However, when we suggested a song they remembered, they were more than happy to play it. Of course, even though one patient needed no help at all (she was a former pianist and had a rather high MMS rating of 22), the others were not able to play a song without extensive help from a musically trained care giver. Furthermore, although 8 or 11 notes are more than enough to improvise, it is hard to find many songs that fit into a single scale.

On the other hand, in the Challenge Mode we tested the following week, patients could play a song without needing help to find the right notes to click on. They clearly were more comfortable following the highlighted notes than trying to create something on their own. We felt we were on the right track because patients would now clap, sing and start reminiscing more and more often. We thus decided to leave the Improvisation Mode as is, and use it only with particularly skilled patients such as the one mentioned above. Interestingly, although this patient was barely able to keep things in memory for more than a couple of minutes, she complained that the playable songs were the same from one week to the other, which greatly surprised the physician among us.

However, in Challenge Mode, failure is rather likely due to the relatively small size of the keys on the screen. Thus, playing the song from start to end demanded quite a lot of time and became rather frustrating without active physical help from a care giver: patients had a lot of trouble following even simple rhythms on their own and were thus unable to recognize the songs. One specific quote illustrates this fact well: when leaving after a session, one patient said "It's hard: we have to remember the songs to sing". This caught our attention because even though this patient had quite a lot of trouble playing in Challenge Mode, what was perceived as hard was not playing, but remembering the song, which was exactly what we had tried to simplify.

We got a lot of similar feedback which showed that patients either (1) are not really aware of how much help with manipulating the Wiimotes they are given when they play or simply (2) do not care about properly pointing and clicking; in any case, they did not seem to mind that the care giver was actually doing almost of the Wiimote handling for them. What they did notice however is whether they were able to remember the melody and lyrics or not. Consequently, we decided to tone down the pointing challenge to an extremely low level and concentrate on the memory aspects of the activity which appeared to have a greater impact on their self-image.

4.3 The Power of Reminiscence

To focus the gameplay on reminiscence, we lowered the game difficulty to a level well below what we had intended in the beginning. First, in Reminiscence Mode (see Figure 3), the highlighted key is three times as big as the others, occupying more than a fourth of the screen, and displays the corresponding lyrics in large letters. Second, instead of having to click in the right place to play, patients can simply hover over the highlighted notes with the trigger pressed in, and the system makes sure that the melody is played with the proper rhythm. One can thus play without even looking at the screen by simply pushing the trigger and shaking the Wiimote from left to right, something that one of our test subjects, even though severely disabled, was thrilled to discover with some practice.

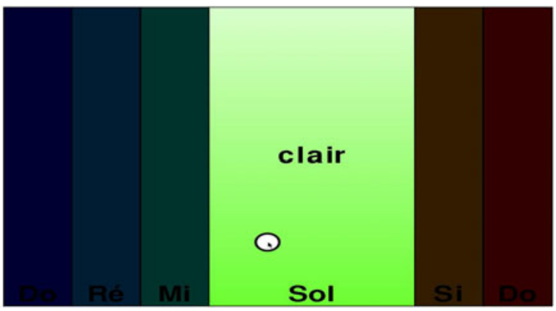


Fig. 3. “Au clair de la lune” in Reminiscence Mode

In fact, we were worried that such an easy task might deter our test subjects by making them feel underestimated. We were completely wrong: none of the patients ever said the task was too easy. Even elderly people without any signs of dementia who happened to pass by during our sessions seemed rather proud to be able to play a song despite the simplicity of the task. We think this is because, for people who were born in the 1920s, any task involving computers is often seen as utterly complicated. Therefore, being able to do even the simplest thing with something so “otherworldly” as a Wiimote is considered an achievement.

This observation is crucial: no matter how easy we make the game, patients will still feel empowered when playing it. We can therefore really tailor our system to trigger as much reminiscence as possible without having to worry about the gameplay becoming dull. This is very important because patients are usually well aware that they forget things more and more rapidly over time and often seem ecstatic when they discover that they do remember the songs from their childhood. Furthermore, the musical memories that come back often trigger a train of other flashbacks (actual quotes: “My husband was a trumpet player;

he often played this when we went to the ball; these were happy times" or "We sung this at X's wedding; how much fun we had back then!").

These recollections usually seem to be even more pleasurable to the patients than the actual singing and can sometimes lead to signs of completely unexpected awareness. For instance, one severely disabled patient (MMS = 10) told us, after playing a song by a famous French singer a few weeks in a row, that she was sad because he was very sick at the time, maybe about to die. The care givers were astonished that a patient with such a high cognitive deficit was able to understand and retain such an information. Thus, even though it could seem like a good idea to design MINWii with the intent of making it playable autonomously, the need to smooth out the impact of such highly emotional occurrences justifies the presence of care givers while the game is being played. Yet, this anecdote also illustrates the tremendous potential of music therapy, which can have unanticipated, very profound cognitive effects [17], far beyond its usually accepted benefits regarding anxiety relief and memory stimulation [13].

4.4 Putting Human Contact First

Our last key design point is human interaction: MINWii is tailored to encourage patients to interact with care givers, family and other patients. Indeed, patients told us praises such as "You're all so nice: it's a pleasure coming here", "Thank you for singing with me: you're a much better singer than I am" or even "For people as lonely as I, human contact is very precious", which led us to reconsider our system not only as a game, but also as a medium for interaction.

First, we never tried to make MINWii usable by patients alone: a care giver always manages the overall progress of the game. For instance, the various songs available need to be presented to the patient in an understandable way. Thus, we decided to implement an accessible file browser with two main characteristics: (1) fonts are very big and all the unnecessary information and buttons are hidden in order not to confuse patients and (2) the titles of the songs are displayed in lieu of their filenames so that patients can read "Oh When The Saints" instead of "ohwhenthesaints.musicxml". This dialog screen is much more important than we had thought at first: demented patients retain their reading ability very long into their illness, to a point that many of our test subjects would read aloud whatever was written on the screen, often without a clue about what it meant. Therefore, just like realizing that they can still remember and sing a song, being able to read and understand written words with the help of the care givers is very pleasing for them, and is a great occasion to interact and give praise.

Second, MINWii's very flexible gameplay, which can be adjusted from quite a serious challenge to an extremely simple task, gives it the ability to appeal to many audiences besides demented patients. More specifically, several informal tests have convinced us that, if children visit, playing with MINWii could be a lot of fun for them and would constitute a very good way to encourage interaction with the patients, mainly through singing and clapping. And anyone who has observed elderly patients can tell that hardly anything could top the pleasure they take in spending some quality time with their great grandchildren.

However, many patients rarely get visits, and the people they see the most are those treated in the same institutions, be it in day-care or long stay units. Just like other people, they chat, take walks and develop friendships, but their illness does make things more difficult, as they have trouble remembering faces and having coherent discussions. Thus, to enhance human interaction, we chose to organize our sessions in the hospital animation room with groups of 3 or 4 patients in front of a big projection screen. Note that, given how easy it is to set MINWii up, it would have been much easier for us, but overall less effective, to plan one-to-one private sessions since demented patients, who have a blurred perception of time, are often quite reluctant to leave their room for fear of missing a visit or a meal, even though nothing is up for them in the next 4 hours.

Our experiment showed that we were indeed able to have patients cooperate in musical production. Since Alzheimer's disease greatly hinders patients' multitasking ability, it is rather hard for the person playing to sing the song at the same time, even though the two tasks are very easy if taken separately. Having group sessions made it possible to still have the patients sing by encouraging an "I play, you sing" dynamic in the group. We were happy to see patients discuss musical preferences, clap and congratulate each other after playing or even, in a few cases, encourage other, more reluctant patients to come to the sessions and play music like "cowboys", as one patient explained because of the pistol.

5 Conclusion

Even though Music Therapy has been shown to be an effective approach to deal with demented patients, few easily usable systems specifically target them. Our MINWii platform, where patients use Wiimotes to either improvise music or play predefined songs on a virtual keyboard, was designed with the specific features of Alzheimer's disease and dementia in mind. Using an iterative process built upon qualitative, small scale experiments in a therapeutic environment, we ended up with a final design that focuses less on creativity but more on reminiscence and renarcissization, which appear to be strong motivators for our test population. A large majority of our patients, with mild to moderate dementia, expressed a strong interest in our system, which was further confirmed by feedback from the care givers, who were able to use MINWii even with no prior musical training.

We are currently undertaking a fully controlled usability study of MINWii, which, if successful, will lead to future research intending to assess its therapeutic impact. A long term goal of our research is also to study the possible use of the interaction data recording facilities of MINWii for applications in evidence-based medicine such as the patient-specific quantitative assessment of Music Therapy compared to other techniques.

Acknowledgments

We thank the patients for their feedback and enthusiasm, Pierrette Despres for her precious help and Nathalie Roy-Sanchez for making this study possible.

References

1. Amyot, A.: Trois cas de figure sur la psychothérapie des personnes âgées. Psychothérapies 1(28), 70–132 (2008)
2. ANAES: Prise en charge non médicamenteuse de la maladie d'Alzheimer et des troubles apparentés (2003), http://www.has-sante.fr/portail/upload/docs/application/pdf/Alzheimer%5Ftechno.pdf
3. Backlund, P., et al.: In: Game Lab, http://www.his.se/ingame
4. Benveniste, S., Jouvelot, P., Lecourt, E., Michel, R.: Designing Wiimprovisation for Mediation in Group Music Therapy with Children Suffering from Behavioral Disorders. In: Proceedings of the ACM Conference IDC 2009, pp. 18–26 (2009), http://www.cri.ensmp.fr/people/sbenven/MAWIIDCPaper.pdf
5. Boulanger, A.: Expressive Gesture Controller for an Individual with Quadriplegia. In: Proceedings of the 10th ACM Conference UbiComp 2008, pp. 113–116 (2008), http://opera.media.mit.edu/Current%20website/Boulanger%5Fubicomp08.pdf
6. Brookmeyer, R., Johnson, E., Ziegler-Graham, K., Arrighi, H.M.: Forecasting the Global Burden of Alzheimer's Disease. Alzheimer's and Dementia 3.3, 186–191 (2007), http://works.bepress.com/rbrookmeyer/23
7. Fluidsynth, http://fluidsynth.resonance.org
8. Folstein, M.-F., Folstein, S.-E., McHugh, P.-R.: "Mini-mental state": A practical Method for Grading the Cognitive State of Patients for the Clinician. Journal of Psychiatric Research 12(3), 198–198 (1975)
9. Guetin, S., et al.: Effect of Music Therapy on Anxiety and Depression in Patients with Alzheimer's Type Dementia: a Randomised, Controlled Study. Dement. Geriatr. Cogn. Disord. 28(1), 36–46 (2009)
10. Guitar Hero, http://www.guitarhero.com
11. Hassenzahl, M.: The Interplay of Beauty, Goodness and Usability in Interactive Products. Human Computer Interaction 19, 319–349 (2004)
12. Lewin, K.: Group Decision and Social Change. Holt, Rinehart and Winston (1958)
13. Livingston, G., et al.: Systematic Review of Psychological Approaches to the Management of Neuropsychiatric Symptoms of Dementia. American Journal of Psychiatry 162(11), 1996–2021 (2005), http://ajp.psychiatryonline.org/cgi/reprint/162/11/1996
14. Non-Invasive Continence Management System, http://people.brunel.ac.uk/%7Eibsreav/NICMS/project.htm
15. Pygame Framework, http://www.pygame.org
16. Riley, P., Alm, N., Newell, A.: An interactive Tool to Promote Musical Creativity in People with Dementia. Computers in Human Behavior 25(3), 599–608 (2009)
17. Sacks, O.: Musicophilia, Tales of Music and the Brain. Vintage, New York (2008)
18. Spector, A., Thorgrimsen, L., Woods, B., Royan, L., Davies, S., Butterworth, M., Orrell, M.: Efficacy of an Evidence-Based Cognitive Stimulation Therapy Program for People with Dementia. The British Journal of Psychiatry 183, 248–254 (2003)
19. Tun, S.-M., et al.: Predictive Validity of Neuropsychiatric Subgroups on Nursing Home Placement and Survival in Patients with Alzheimer's Disease. Am. J. Geriatr. Psychiatry 15(4), 314–327 (2007)
20. Vas, C., et al.: Alzheimer's Disease, The Brain Killer. World Health Organization manual (2001), http://www.searo.who.int/LinkFiles/Health%5Fand%5FBehaviour%5Falzheimers.pdf
21. Yam, M.: Wii Invades Retirement Home, http://www.dailytech.com/Wii+Invades+Retirement+Home/article6191.htm

Virtual Team Performance Depends on Distributed Leadership

Nico van Dijk and Joost Broekens

Media Technology MSc, Leiden University
Man-Machine Interaction Group, Delft University of Technology
nicovandijk@gmail.com, joost.broekens@gmail.com

Abstract. In this paper we present a detailed analysis of World of Warcraft virtual team collaboration. A number of competitive synchronous virtual teams were investigated in-situ and unobtrusively. We observed a large gap in team performance between the various teams. An initial statistic study showed that, in teams of this level, individual player performance was not the primary driver for the large discrepancy in team performance. This led to the argument that differences in intra-team collaboration and communication might be a significant driver for the discrepancy in team performance. In total 16 hours of audio recordings of gaming sessions of virtual teams were analyzed. The analysis indicates that distributed leadership instead of authoritative leadership is more common in successful synchronous virtual teams.

Keywords: Virtual Teams, Collaboration, Communication, Distributed Leadership, MMORPG.

1 Introduction

Recently, the use of computer games for research purposes has become increasingly popular [1]. Often this research is focused on researching how games can be made to be more entertaining or educational, but also on how video games can be used as research tools to study more general phenomenon (see, e.g., [2, 3]). Topics such as team strategy and decision making are being researched in a video game context [4].

Virtual teams have become more prevalent in various fields of business and research as Computer Mediated Communication (CMC) has become more widespread and accessible. A recent survey by the Gartner group found that more than 60% of professional employees work in virtual teams [5, 6, 7]. It is expected that the business world will increasingly use CMC [6]. Additionally, a significant amount of study on virtual teams and team training in virtual worlds is being done in defense and disaster relief related areas [8, 9]. Thus it is very valuable to investigate the internal communication of teams already operating under these conditions.

There are still gaps in knowledge about virtual team collaboration and communication, in particular when it comes to studying virtual teams in-situ. Therefore it is worthwhile to study teams with a real-time, unobtrusive method that can be practically implemented [10, 11].

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 91–102, 2010.

In online computer games, especially Massively Multiplayer Online games (MMO), CMC is the rule rather than the exception and this creates interesting opportunities to study Virtual Teams in-situ. MMO's serve as particularly good environments for this type of research for a number of reasons. MMO's have active functioning teams that share a common task and goal. Due to the amount of players there is a large diversity of teams available to collect data on. The members of these teams are self-motivated to participate in their shared activity. Members can be observed unobtrusively and in real-time while being engaged in sessions of collaborative problem solving. Players invest quite some effort in the game and their team indicating that the team is indeed viewed by the players as an online team with a common underlying interest. Finally, player behavior with respect to leadership and other higher level organizational skills is quite consistent with respect to the player's real-world behavior (see e.g. [12]). We studied communication and organization behavior in synchronous virtual teams in the popular MMO World of Warcraft (WoW).

We set out to investigate the relation between team performance and team structure. More precisely, we wanted to know if there was a relation between the quality of the team and the type of collaboration the team used. The study setup was unique in that the participating teams operated in duplicate environments, while still being in-situ and unobtrusive. This means that variation in team behavior is caused by the players and their interactions. The teams were unaware of the study, providing an excellent opportunity to study the teams in the wild. The data was analyzed qualitatively. This paper is organized as follows. In section 2 the relation between virtual teams and MMO's is explained. In section 3 our study setup is explained. In section 4 the results are presented. Section 5 and 6 present the discussion and conclusion respectively.

2 Virtual Teams and MMO's

Cohen and Baily define a team as *"a collection of individuals who are interdependent in their tasks, who share responsibility for outcomes, who see themselves and who are seen by others as an intact social entity embedded in one or more larger social systems, and who manage their relationship across organizational boundaries."*[13] This definition is general enough to capture traditional, as well as virtual teams while precisely identifying the defining features of a team: its unity of purpose, its identity as a social structure, and its members' shared responsibility for outcomes.

Rosen et. al define a virtual team as a *"group of geographically, organizationally and/or time dispersed workers brought together by information and telecommunication technologies to accomplish one or more organizational tasks"* [14, 15, 16].

2.1 Virtual Teams in World of Warcraft

The Virtual teams that were studied in WoW adhere to the requirements of both of these definitions. Players form large collectives called *guilds*. From these guilds smaller teams of players are formed into what is known as a 25-player *raid group*. We consider a 25-player raid group as a team (please note that as all guilds in our study had one 25-player raid group, the terms guild, team and 25-player raid group will be used interchangeable and all refer to the guild's 25-player raid group). This 25-player raid group attempts to succeed in the bigger challenges that can be found inside the game environment. The members of these groups are interdependent in their tasks, communicate only through computer mediated means and are driven to succeed in collectively overcoming the challenges that the game offers. Considering the time investment that is needed from the team to defeat these in-game challenges (typically more than 20 hours of play time) and the amount of communication occurring we assume that these challenges are complex problems. While their livelihood does not depend on their success in the game the amount of time team members dedicate to their team is significant enough that we can consider the team to generally be very well motivated to succeed. There are many similarities between our virtual teams and the problem characteristics of virtual teams described by the Macro-cognitive Model of Team Collaboration [17]. Virtual teams in WoW operate under time pressure and often not all information or knowledge about a challenge is available for players. There usually is a large amount of knowledge to handle and the interface with the game is complex. Player's often custom design their user interface to facilitate their functioning in the team. The teams are distributed, culturally diverse, usually employ a fairly strict--although not very deep—hierarchy, and their members have unique roles and regularly rotate within the team. The operational tasks are also similar, team decisions about the planned tactics have to be made, the players have to develop a shared understanding of the situation and transform available information into a functional tactic that works for the team in the field. One difference is that the teams tend to work synchronously instead of asynchronous. Asynchronous communication usually only occurs during the planning phase for content when the teams use forums to exchange information.

2.2 Distributed and Authoritative Leadership

The following two descriptions for collaboration styles were based on phenomenon found in the data initially. When a team employs authoritative leadership, there is a designated person that functions as leader. Communication is characterized by giving instructions, explaining decisions and asking about status information (i.e., information important for decisions, planning and action selection is typically gathered and analyzed by the leader). The leader decides what the team does.

In contrast, in distributed leadership, leading never rests with just a single person but rather leadership is a role often and quickly passed between different members of a team. Teams that use distributed leadership are characterized by high communicative interaction between its various members. Focus is shifted away from the individual actions of leaders and their personal qualities. Instead the practice of leadership as "interactions between the leaders, followers and situation" becomes more important, and this interaction is used to plan, decide and act [8].

3 Research Design

We now explain our research design used to investigate the relation between team performance and team structure. The chosen gaming platform enabled us to study intra-team communication in great detail: all data can be logged, and the researcher's real motives could be hidden. These two aspects have not often been exploited in similar studies. Given the large amount and variability of data collected, and the relatively open-endedness of a participatory field study, a qualitative study design was more suitable as compared to a quantitative hypothesis-focused approach. This allowed our data analysis to be more explorative.

It is plausible to assume that differences in team performance are explained by individual player performance (i.e., better players in better teams). We carefully checked this assumption before starting the detailed and time consuming task of monitoring and qualitatively analyzing teams in order to find effects of team collaboration on performance. We polled a random set of 150 players in the top 4000 guilds in Europe (top 4000 selected according to www.wowprogress.com, 5 players per guild, 30 guilds in total). These players were ranked according to a measure for player quality based on several game-performance elements extracted from the website europe.wowarmory.com (official website of the developer Blizzard containing actual game data). These elements included: number of bosses killed in 5 player dungeons, number of 5 player dungeons entered, number of bosses killed in 10 player dungeons, number of 10 player instances entered, number of Quests completed, number of Daily Quests completed, date of reaching level 80.

In essence, all elements define a player's activity in the game, except the last which defines how "senior" a player is with respect to how long the player plays the game. We excluded the activity related to the team performance we will be observing (i.e. the 25-raid group activities). Players with high activity and seniority have more experience and are assumed to be better performing.

To investigate if player rank relates to team rank in the top 4000 teams, we statistically analyzed this relation. This showed a significant correlation between guild rank and player rank among the top 4000 teams ($n=150$, $r=-0.273$. $p=0.001$). However, when considering only the teams ranked 1000 to 4000, there was no significant correlation ($n=99$, $r=-0.127$, $p=ns$). This indicates that within the well-performing set of ranked teams, not belonging to the absolute top 1000, individual player quality is not related to team performance. Thus, in the top 1000 to 4000 other factors, such as collaboration and communication explain the difference in team rank.

Based on these results we decided to embed an avatar (representation of the researcher in the game) in European guilds ranked between position 1000 and 4000. These avatars would be active in the guild's 25-raid group for several weeks / months. The main goal for our study design was to obtain all our data unobtrusively and in-situ. As such, the fact that research was being done was hidden to the virtual team. In ethnographic fieldwork it is quite accepted to study subjects without formally requesting their consent. We consider our method of acquiring data to share many similarities with this process.

A secondary goal was to gather data on teams during encounters they had not defeated before. Prior experience by the first author indicated that a team's communication pattern changes drastically when teams have defeated the same challenge several times. To eliminate this factor only data recorded during moments teams were engaging new content was analyzed.

Four teams to gather data from were selected based on guild rank (according to www.wowprogress.com and our previous analysis), and if the guild had already finished a newly published challenge (teams attempted similar new content). As clarified in 2.1 we use the term guild and team interchangeably and they refer to the same entity.

- Team A – position ~1000
- Team B – position ~1300-1500
- Team C – position ~1500-1800
- Team D – position ~4000

The differences in performance between guilds ranked near position 1000 and guilds ranked near position 4000 are significant. For example, Guild 'Going Postal' ranked EU 1025 achieved the 'Heroic Lord Jaraxxus' challenge on October 3rd 2009, while Guild 'Ypsilon' ranked EU 3747 achieved the same challenge on December 21st 2009. This represents a difference of more than two months, between guilds that both play on 3 scheduled evenings each week. The boss became accessible to both teams in the first week of September 2009, meaning 'Going Postal' was close to 400% faster, given that they played on similar schedules. This is a major difference given the fact that these guilds are ranked and being the best team and achieving a higher rank on the server or continent is an important status achievement in WoW.

Data was gathered in several different ways. WoW raid groups use Voice Over Internet Protocol (VOIP) applications during their team sessions to facilitate their intra-team communication (e.g. Teamspeak or Ventrilo). These voice communications were recorded. Audio recordings were started the moment a group was fully gathered and players got on their way to the respective challenge they would attempt that evening. Audio recordings were concluded when the group was disbanded at the end of the evening.

3.1 Data Analysis

Most interesting were the VOIP audio recordings, these seemed to contain the highest frequency of communication. Nivo8 was used to structure the data. The data analysis was explorative and a coding scheme was developed to disseminate the communication. After initial scanning, it seemed necessary to develop a more structured and generalized coding scheme that could be applied to both the audio recordings and the chat logs as well as be generic enough to be able to be applied to virtual teams in different environments.

All coding of the audio files was done manually. This was a very time intensive process. In total 16 hours of audio data (out of the total 70+ hours recorded) was analyzed. Session length was between 50 minutes and 4 hours long. Every session was coded and annotated down to the second. Six Sessions in total were analyzed, two for Team A, two for Team D, one for Team B and one for Team C. The number of references being created per session varied between approximately 700 and 2800 depending on the length of the session.

To code what happened in the audio and chat, we used two coding schemes. These schemes were able to cover all information exchange within the teams that was related to the challenge they were attempting to succeed in.

- Members: (Officer1, Officer2, Officer3, Officer4, Officer5, Raiders)
- Tactical information: (Instruction, Negotiation, Reflection, Status, Action)

The *member* scheme represents who speaks. All teams that were studied had defined officers and the remaining members of the team were called raiders. It's worthwhile to mention that these guilds determine and assign their own officers inside the game. They are regular players, but in the guild have extra privileges and control over who is member in the guild, as well as carry extra tasks and responsibilities (administration, recruiting, conflict management etc.). This self-regulated hierarchy was used to assign members to their respective codes. Raid groups usually consisted of 2-4 officers and 21 to 23 raiders (in total 25 members).

The *tactical information* scheme contains all audio fragments that were related to sharing information, making decisions and forming a strategy that was concerned with the content that was being done by the team during that session. In order to determine what code could be applied to an audio fragment several semantic rules were used.

Instruction: Sentences that implied that a different member had to perform an action, or perform an action in a certain way. "Do x", "watch out for y", "perform z 20 seconds into the fight". This code contains both overarching instructions, as well as direct warnings for certain occurrences in the game environment.

Negotiation: Any discussion about the current strategy that the team was using that involved one or more members. "Should we not do x in manner z instead of manner y", "No, manner z is superior to manner y because of factor a". This code contains suggestions, more in-depth discussions, and consensus reaching.

Reflection: Any audio fragments that were concerned with establishing what happened in the past. Whether it was about judging what went wrong for the team during the last attempt, or in order to establish what happened in a similar situation several days before.

Status: Any audio fragments that were concerned with players reporting to the group about their current condition, the state of the environment or the state of the group.

Action: Any audio fragments that were concerned with declarations of action and intention. "I'm going to do x".

4 Results

Two sessions were analyzed for both Team A and Team D to eliminate the chance of a fluke result. The Member scheme results (Fig. 1) indicate that Team A had several leaders that were equally active in their communication. The remaining 21 players that are coded under the Raider label formed the largest share of the communication over the course of the sessions. 35.9% of all the audio recorded during the first session (Team A session 1, TAs1) was spoken by the regular raid members. During this session three of the four officers spoke almost an equal amount. Officer1 24.0% of the time. Officers 2, 3 and 4 we're responsible for 17.9%, 19.3% and 3.0% of the audio respectively. It also seemed that the officers had a rough task distribution. Two officers (Officer1 and Officer3) were primarily concerned with explaining tactics, and one of these in particular initiated many reflective conversations after the team made an attempt at the boss they were trying to defeat. The second officer (Officer2) was primarily concerned with logistics, and made sure all team members were present, ready and had all their necessary avatar enhancements. Whenever an attempt was started he was always the person to give the go signal for the team to initiate the fight. The second session recorded for Team A (TAs2) was recorded during a slightly more

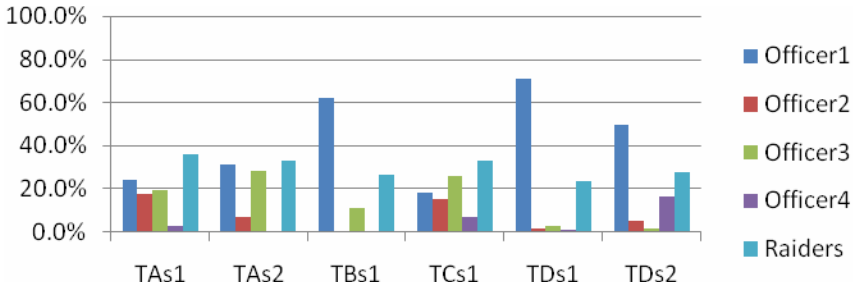


Fig. 1. Team member audio activity

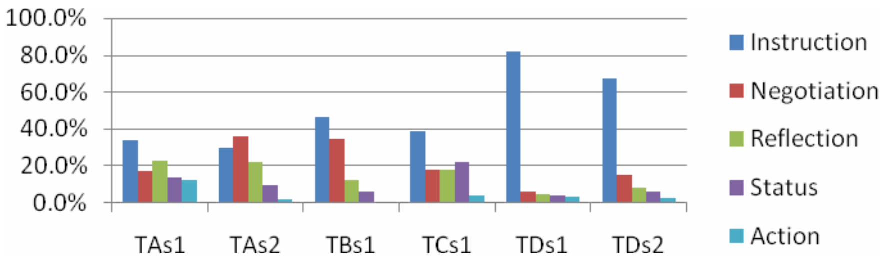


Fig. 2. Team communication type coverage. Notice the distribution among both of Team A's logs compared to both of team D's logs.

complex engagement. The team had quite some difficulties with the challenge and many attempts were made that usually ended fairly quickly. Officers 1 and 3 primarily concerned themselves with the tactics and spent a lot of time discussing among each other and with the rest of the team. Officers 1 and 2 were responsible for 31.5% and 28.5% of the audio respectively The second officer (6.8%) was not in control of team logistics that session and was much less active in the discussions. The group of Raiders again formed the most active communication participant (33.3%).

If we review Fig. 2 for Team A it is noticeable that all types of communication (instruction, negotiation, reflection, status and action) are all well represented in both sessions. Significant amounts of time were spent negotiating tactics, reaching team consensus and reflecting on past attempts. The first session contained 34.2% Instruction, 16.9% Negotiation, 23.1% Reflection, 13.6% Status and 12.1% Action messages. During Team A's second session negotiation was even the most populated code. The second session contained 30.1% Instruction, 22.2% Reflection, 36.3% Negotiation, 9.4% Status and 1.9% Action messages. In all other sessions, instruction was the most populated code. The problems encountered by Team A in this second session might explain their increased need to negotiate about a working tactic on the spot.

Team D had a very distinctly different communication pattern. Both of the sessions for Team D were recorded on the exact same fight as the first session for Team A (TAs1). Team D was controlled by a very directive officer (Officer1). On his own he tended to speak at least twice as much as the remaining 21 members of the group. The larger share of the group showed very passive behavior. The first session for Team D can be broken down as follows: Officer1 71.1%, Officer2 1.4%, Officer3 2.6%, Officer4 1.2%, Raiders 23.7%. Where Team A used four officers to handle the task workload of leading, Team D also had four present but only one of them was very communicative. The second session for Team D showed very similar results to the first session: Officer1 49.8%, Officer2 4.9%, Officer3 1.7%, Officer4 16.1%, Raiders 27.5%. This similarity is also found in the coded references of the *tactical information* scheme. The gap between Instruction and the other communication forms is very large in the data of Team D. The first officer was also the primary source for these Instructions. The *tactical information* scheme analysis for the first session contained 82.3% Instruction compared to 4.3%, 6.0%, 4.1% and 3.3% for Reflection, Negotiation, Status and Action respectively. The second session contained Instruction 67.4%, Reflection 8.4%, Negotiation 15.0%, Status 6.4%, and Action 2.9%.

Let us compare a particular section of the data where distinct communication style differences between Team A and Team D occurred. We'll take a look at the same situation in logs TAs1 and TDs1, during which both teams were attempting to defeat the challenge 'Beasts of Northrend, Heroic'. During the first section of this fight the tanks[1] had to rotate among them who was tanking the boss during the encounter based on what number of debuffs[2] they had received from the boss. In Team A the tanks declared on their own how many debuffs they had, using a predetermined rotation order the tanks knew at what number of debuffs they had to take over and in which order they had to take over. These declarations of debuffs were coded as Status messages. Whenever a tank reached the number that was the trigger for a rotation to occur he would simply say the number over the VOIP channel and that would signal to the next tank in the rotation to take over. The tank that was taking over would often declare his action before taking over from the first tank; 'I'm about to take over' in order to inform the healers[3] assigned to keeping the tanks alive to switch targets to the tank that was stepping in. These messages were coded as Action messages. This encounter was unforgiving enough that if a tank went without any healing for a period of 3 seconds he would very likely be dead. A death on the part of one of the tanks generally resulted in a failure for the entire team. This made the well coordinated switching of the tanks and healers very important.

In Team D this entire communication process was handled by the first officer. Whenever a tank reached the maximum amount of debuffs the officer would order one of other tanks to take over and would also instruct the healers to heal the tank that was stepping in. This put a lot of stress on the first officer and also led to create confusion among the tank rotation as well as the healers. Whenever the officer used a loud voice to instruct the healer to heal one of the tanks, healers with assignments other than healing the tanks would drop their original assignment and switch targets, which would then lead to increased risk to their original assignments. This deficient communication was the direct cause for several failed attempts and also created a very uncomfortable atmosphere in the group that caused the fight to be perceived to be harder than it actually was if the communication had been more distributed and

[1] Tanks: a particular role in the team filled by several members, their job is to protect the weaker members in the team by drawing the attention of threats to themselves.

[2] Debuffs: a temporary negative effect put on an avatar.

[3] Healers: players tasked with keeping other team members alive.

calmer. The lack of reflection and negotiation during the sessions of Team D (Fig. 4.2) also indicated they failed to improve their tactics during the evening and were not able to determine what went wrong during attempts to make an improvement. Team A reached the 3rd of this 3 phase fight near the end of the evening while Team D only managed to reach the 2nd phase of the fight on their last attempt of the evening.

The data from Team A suggested that the ability to clearly define errors and problems were a critical part for their success. Defining where mistakes were made in the team and clearly declaring them made the team much more efficient and also created a more confident atmosphere. When the team wasn't succeeding they always knew why they were not succeeding. Among Team D this did not occur as much. Often the same set of instructions would get repeated even when no improvement was observed.

In order to verify if the found communication patterns can be extrapolated we analyzed the data of Team B (~1300-1500) and C(~1500-1800) (both teams are positioned between Team A and D with respect to performance,. The session analysis for Team B resulted in Officer1 62.0%, Officer3 11.3% and Raiders 26.7%. Team B shared similarities with Team D in that they had a very authoritative leader. However, their *tactical information* scheme analysis shows a high amount of negotiation. The tactical analysis for Team B can be broken down as follows: Instruction 46.1%, Reflection 12.4%, Negotiation 34.6%, Status 6.2% and Action 0.7%. So, while Team B shared similarities with Team D's communication pattern, Team B was not nearly as instructional or authoritative. The session analysis for the slightly lower ranking Team C produced Officer1 18.5%, Officer2 15.4%, Officer3 26.1%, Officer4 6.9% and Raiders 33.2%. They shared a similarly distribution of workload as Team A, but the number of references (almost twice as much references per time unit compared to Team A) coded for their session indicated that their communication was very fragmented and that members spoke often for very short lengths. This was reinforced by the results of the *tactical information* scheme. Team C's breakdown resulted in 38.5% Instruction, 4.3% Reflection, 6.0% Negotiation, 22.2% Status, 4.0% Action. Their number of Instruction references shared more similarities with Team D's *tactical information* scheme results.

5 Discussion

Very distinct differences were found between the communication patterns of the highest ranked team (position ~1000), Team A, and the lowest ranked team (position ~4000), Team D, in both of the sessions of each team. Large differences were evident in the *member* scheme (Fig 4.1) as well as in the *tactical information* scheme (Fig. 4.2). The higher ranked Team A showed a distributed communication pattern where communicative workload was equally divided between its leaders, and where the group of raiders, that numbered 21 members, was always the most vocal contributor to the total communication. The type of communication that was observed was also equally distributed; Team A showed high amounts of reflection and negotiation during their sessions. Because of the performance record of Team A we argue that this style of communication is an important asset of their team's success.

The challenges these teams face in WoW can be described as a 25 entity collaborative reinforcement learning problem. Teams tend to spend dozens of attempts and many hours learning how to overcome a particular challenge while slight improvements are made with each attempt. We assess that it is very important for a team to have all its members learn and adjust to a new piece of content as quickly as possible. The team that can enable its members to learn fast ends up being successful, and we suspect that higher performing teams manage to achieve this through high interaction between all members of the team and very consciously negotiating and reflecting on the tactic they are developing to overcome the challenge they are facing. Interestingly enough, while the number of scheduled evenings per week and player level for Team D is roughly equal to Team A, their communication style was almost the exact opposite. Team D's communication was characterized by a single dominant leader who had the tendency to micromanage the team members actions. The remaining 24 members of the team were not nearly as communicative as this single leader and this is also well reflected in the communication type analysis.

Instructions were by far the most common, often coming from the dominant leader, but when other members of the team did speak up it was also often in the form of an instruction. The high amount of instructions coupled with a lack of reflection and negotiation generally meant the team tended to stick to their original plan and would not revise it very often after learning more about the content. This inflexibility and the low interaction between the various members led to a slow learning curve for the team. The results from Team B and C interpolated well between the results from A and D. In our study setup Team C was ranked lower than Team B, despite sharing slightly more similarities with Team A rather than Team D. However due to the way content was released and the time it took to record the necessary data the ranking difference between Team B and Team C is not as straightforward as it may seem. The guild ranking system is quite dynamic and new achievements by guilds quickly lower or raise a team's standing with about 100-300 points.

Our results replicate the results of other team related studies. In a closed experiment with novice users by Richter and Lechner, the most successful team was also the team using negotiation and reflective analysis the most [18]. However no data was available on the distribution of communication of each member of the 5 man team. A larger study of Change Management Teams also showed shared leadership as a more effective predictor of higher performing teams, however their research methods relied on two questionnaires applied at the start and end of a 6 month period[19].

Certain types of communication have different effects on how fast they contribute to the construction of a team's shared mental model [20]. Based on our results we suspect that negotiation and reflection contribute more effectively to shared mental model development than other communication types like instruction. The faster a team can switch from explicit communication to primarily implicit communication, when they are operating in an environment with a lot of time pressure, the faster it leads to a successful collaboration. Especially members of WoW raid groups benefit a lot from implicit communication as the act of communicating tends to negatively affect their ability to perform their in-game function during the team effort.

6 Conclusion

Distributed leadership is a method of collaboration that involves multiple leaders that delegate and switch roles to each other and to non-leaders in order to facilitate the performance of the team. According to the results we gained from gathering data on four WoW teams in European top 1000 to 4000, the better teams tend to use distributed leadership. The collaboration among the better teams in our study is marked by a high interaction between both leaders and followers and team strategy development with high levels of reflection and negotiation. The slightly lower ranked teams B and C use a somewhat distributed communication style. In contrast, the lower ranked team's collaboration (D) is characterized by a rigid command structure involving instruction and control of a dominant single leader. This suggests that the top teams in WoW are more likely to be using distributed leadership as opposed to a more top-down instructional approach that more commonly occurs in lower ranked teams. This means that there is a relation between the distributiveness of a team's communication pattern and the team's performance. As these processes are general aspects of team collaboration, and as other studies have found similar results in different environments (e.g. [19]), we argue that these findings can also be applied to other synchronous virtual teams that rely on a high information exchange in order to succeed, especially in areas like disaster relief and defense related team efforts. Communication in these teams is essentially information exchange. Learning which communication style is more prevalent in more successful teams can provide important knowledge on more effective information exchange & consensus reaching within virtual teams in general. Considering the novelty of the study method and the platform, it is an encouraging result for further team related studies in this area.

References

1. CHI Workshop on video games as research instruments, `http://www.uclic.ucl.ac.uk/people/e.calvillo/ws_videogames/`
2. Warner, N., Letsky, M., Cowen, M.: Cognitive model of team collaboration: Macro-cognitive focus. Human Factors and Ergonomics Society Annual, USA (2005)
3. Broekens, J.: MMOGs as Social Experiments: the Case of Environmental Laws (2008) arXiv:0811.0709
4. Dannecker, A., Richter, S., Lechner, U., Dressner, N., Fabisch, S., Ilsemann, A.: Towards World of Warcraft as an experiment platform for teams. In: Americas Conference on Information Systems, USA (2008)
5. Martins, L.L., Gilson, L.L., Maynard, M.T.: Virtual Teams: What Do We Know and Where Do We Go From Here? Journal of Management, 805–835 (2004)
6. Kanawattanachai, P., Yoo, Y.: Dynamic nature of trust in virtual teams. Journal of Strategic Information Systems 2(2), 42–58 (2002)
7. de Lisser, E.: Update on small business: Firms with virtual environments appeal to workers. Wall Street Journal, B2 (1999)
8. Reeves, B., Malone, T., O'Driscoll, T.: Leadership's Online Labs. Harvard business review, USA (2008)
9. Jensen, J.A.: Joint Tactics, Techniques, and Procedures for Virtual Teams. USCINCPAC (2002)

10. Rosen, M.A., Fiore, M.S., Salas, E., Letsky, M.: Tightly Coupling Cognition: Understanding How Communication and Awareness Drive Coordination in Teams. The international C2 Journal, USA 2(1) (2008)
11. Salas, E., Cooke, J., Rosen, M.A.: On Teams, Teamwork, and Team Performance: Discoveries and Developments. Human Factors 50(3), 540–547 (2008)
12. Powell, A., Piccoli, G., Ives, B.: Virtual Teams: A Review of Current Literature and Directions for Future Research. The DATA BASE for Advances in Information Systems, USA 35(1) (2004)
13. Reeves, B., Malone, T.: Leadership in Games and at Work: Implications for the Enterprise of Massively Multiplayer Online Role-Playing Games. Seriosity, Inc., Palo Alto (2007)
14. Cohen, S.G., Bailey, D.E.: What makes teams work: Group effectiveness research from the shop floor to the executive suite. Journal of management 23(3), 239–290 (1997)
15. Alavi, M., Yoo, Y.: Is Learning in Virtual Teams Real? Harvard, Boston (1997)
16. DeSanctis, G., Poole, M.S.: Transitions in teamwork in new organizational forms. Advances in group processes 14, 157–176 (1997)
17. Jarvenpaa, S.L., Leinder, D.E.: Communication and trust in global virtual teams. Organization science 10(6), 791–815 (1999)
18. Richter, S., Lechner, U.: Transactive Memory Systems and Shared Situation Awareness: a World of Warcraft experiment. In: International Conference on Organizational Learning, Knowledge and Capabilities, The Netherlands (2009)
19. Pearce, C.L., Sims Jr., H.P.: Vertical Versus Shared Leadership as Predictors of the Effectiveness of Change Management Teams: An Examination of Aversive, Directive, Transactional, Transformational, and Empowering Leader Behaviors. Group Dynamics: Theory, Research, and Practice 6(2) 172–197, (2002)
20. Mathieu, J.E., Heffner, T.S., Goodwin, G.F., Cannon-Bowers, J.A., Salas, E.: The Influence of Shared Mental Models on Team Process and Performance. Journal of Applied Psychology 85(2), 273–283 (2000)

Nonverbal Behavior Observation: Collaborative Gaming Method for Prediction of Conflicts during Long-Term Missions

Natalia Voynarovskaya, Roman Gorbunov, Emilia Barakova, Rene Ahn, and Matthias Rauterberg

Eindhoven University of Technology,
Postbus 513, 5600MB, Eindhoven, The Netherlands
{n.voynarovskaya,r.gorbunov,e.i.barakova,r.m.c.ahn}@tue.nl,
g.w.m.rauterberg@tue.nl

Abstract. This paper presents a method for monitoring mental state of small isolated crews during long-term missions (such as space mission, polar expeditions, submarine crews, meteorological stations, and etc). It combines the records of negotiation game with monitoring of the nonverbal behavior of the players. We analyze the records of negotiation game that has taken place between the crew members who were placed in isolated environment for 105 days during the Mars-500 experiment. The outcomes of the analysis, differently from the previously made conclusions, show that there was not a significant deviation of the rational choice of the players. We propose an extension of the method that includes monitoring of the nonverbal behavior of the players next to recording the game records. The method is focused on those aspects of psychological and sociological states that are crucial for the performance of the crew. In particular, we focus on measuring of emotional stress, initial signs of conflicts, trust, and ability to collaborate.

Keywords: Colored Trails, social network analysis, nonverbal communication, emotions, long-term missions, evolutionary game theory.

1 Introduction

Long-term missions, such as a space flight or polar expedition, usually are performed by small group of people, and characterized by extreme physical and psychological parameters. The psychological parameters, like small group living, extreme social isolation, working together in close proximity [2], could be crucial for the mental state of crew members as well as the whole crew. The interpersonal issues, such as crew tension, cohesion, leadership, language, cultural factors [1], could become reasons for personal conflicts which, in their turn, could affect performance of the mission. For instance, inter-crew tension can cause formation of subgroups, disruptions of cohesion, scapegoating, refusal of communication and collaboration.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 103–114, 2010.

For measuring, prediction and prevention of the above mentioned problems, it is crucial to have methods for automatic monitoring of psychological and social states in the group. The psychological states of the crew members, as well as different aspects of interpersonal relations, have been assessed using different approaches. In the current study we combine three approaches: strategic multiplayer game, [8, 11], techniques for automatic monitoring of nonverbal behavior [16,17,18,21,22,23], and self assessment based on questionnaires [12,13,14,15]. In this way we are aiming to overcome limitation of every component and develop a self consistent and comprehensive technique for monitoring intra- and inter-personal state of the crew members.

2 Related Work

The first concept of the AMHA project has been done within Mars-500 experiment for 105 days isolation [4, 5, 6, 7]. In 2004 the Institute for Biomedical Problems (IBMP) in Moscow and the European Space Agency have started to plan a full-scale ground based simulation of a manned mission to Mars. Such a full scale mission requires between 520 to 700 days of isolation. Referring to the lower end of this time frame the initiative was named Mars-500.

The goal of the Mars-500 study is to gather data, knowledge and expertise required to prepare a real mission to Mars. Hence, all key peculiarities expected to be present during future missions to Mars are reflected: ultra-long duration, need for autonomy, affected communication due to signal delay, and limited stock of expendables.

This ensures that psychological and physiological impacts of isolation through such an extended period of time are observed as close to reality as possible. A crew of six candidates (four Russians and two from EU countries) are sealed insight the facilities of the Institute for Biomedical Problems in Moscow. An initial 105-day isolation period took a place in spring of 2009. The present concept is going to be tested in the full 520-day study, which has launched at the 3rd of June, 2010.

The Mars-105 experiment presents the crew members encountered in strategic game interactions while data is being gathered about their interpersonal dynamics.

2.1 Evolutionary Game Theory and Colored Trails Game

In this project, we explore the use of strategic multi-player games to alleviate stress, and more importantly as an unobtrusive tool to monitor the mental capacity of astronauts as well as the development of different social interaction patterns within the crew. We are primarily interested in games that feature the following properties:

- Simple enough for analysis
- Rich enough to reflect features of real life interactions
- Grounded in a situated task domain
- Strategic (i.e. partial information that promotes reasoning)
- Suited to measure social factors such as fairness

We are also interesting to produce data, which directly involve interpersonal relation and, as a consequence, can be interesting to monitoring social atmosphere in the crew.

As such a tool, we used a three-player negotiation variation [25] of the Colored Trails framework developed at Havard University [24]. The Colored Trails game is played on a board of colored squares. One square is designed as a "goal square". Each player has a piece on the board and possesses a set of chips in colors chosen from the same palette as the squares of the board. To move his own piece into an adjacent square a player must turn in a chip of the same color as the square. Chips can be exchanged by the players if a mutual agreement is reached. The goal of the game is to move your own piece as close to the goal square as possible using as less chips as possible. Distance to the goal is more important than number of chips left after the move.

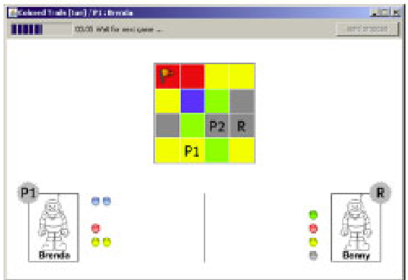


Fig. 1. Colored Trails Game

2.2 Data Analysis of Mars-105 Experiment

Design of the presented monitoring system is based on our previous analysis of the data obtained in the MARS-105 experiment. In particular behavior of Responders was analyzed. Situation, in which a responder needed to make a decision, was described by two numbers: how the Responder can improve his/her payoff by accepting the first offer and how he/she can improve his payoff by accepting the second offer. This description of the situation is a simplification because other factors can influence the decision of the Responder (for example what other players will get as a result of his/her decision or how good are the available offers in comparison with the offers which could be proposed). However, the two selected parameters are the only factors, which are determining behavior of the Responder and showing if the player's behavior is totally rational. So, the mentioned two parameters can be used to check how rational the behavior of Responder is. This check is important because any deviation from the rational behavior can potentially be an indication of psychological preferences or other interpersonal relations.

On the Figure 2 is shown the behavior of one of the Responders. The x- and y-axis correspond to the improvement of the payoff which can be achieved by accepting the first and second offers, respectively. So, generally speaking, every point in the plot represents a particular situation in which the given Responder needed to make a choice. The color of the point indicates the choice which was made. The red/blue colors mean that the first/second offer was accepted. The green color means that the both offers were rejected.

The space on the graph is divided into three different regions. If a responder is totally rational, every region should only contain points of one corresponding color.

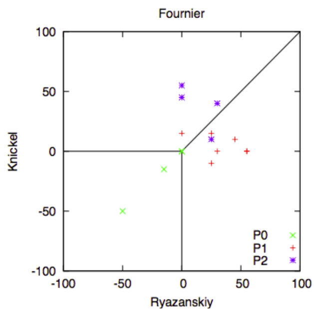


Fig. 2. Dependence of the behavior of one Responder on the accepted offer

For examples, if both offers worsen the payoff of the responder (the left bottom region) than both offers should be rejected (green points). The right-bottom region means that the first offer is better than the second one and it improves the payoff of the responder (so, in this case the first offer should be accepted (red points)). In the third region (left-top) the second offer should be accepted (blue points).

3 Method

3.1 Design Concept

Based on the previous AMHA design, the study of the Colored Trails game performed during MARS-105 [4, 5, 6, 7] and data analysis, presented in previous section and [4], we lay down conditions and requirements to extend the rules of the game, and experiment by itself.

To build the AMHA concept model, firs of all, we defined the parameters, which are crucial for successful communication and collaboration, such as stress level, pleasure, trust, dominance, and etc. Furthermore, we defined a personal mental state and interpersonal relationships as two main components of the crew mental health. The combination of different techniques is required to assess different aspects of the psychosocial states as well as to perform cross validation and correct interpretation of the collected data (see Figure 3).

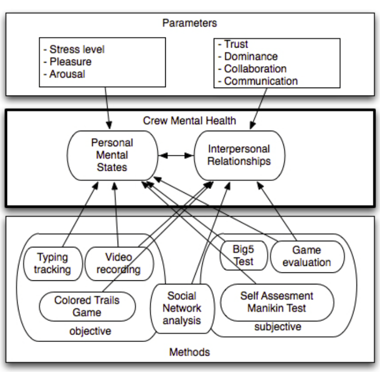


Fig. 3. AMHA conceptual model

Usage of questionnaires for a self assessment is important component of our design, since crew members can provide explicit information about their psychological states and interpersonal relations in the crew. This information is important for interpretation of the game results in terms of perception of psychosocial state of the crew.

3.2 Generalization of the Colored Trails Game

For the MARS-500 experiment we have proposed a generalization of the Colored Trails game. In the new version of the game every user plays as Proposer. Moreover, in contrast to the previous version of the game, Proposers are free to choose a player (Responder) whom he/she wants to offer a chips exchange. As a consequence of these modifications, two or three players will play the Responder role in the second stage of the game.

The game flow has four phases.

1. Choosing a partner
   In this phase each player is choosing a partner for the current session.
2. Proposition phase
   In this phase, each player, acting as a Proposer, can offer a chips exchange to his adversary. The Proposer's and his Adversary chips are always located on the left and right hand side of the window, respectively. On the same screen the player can see the game overview.
3. Response Phase
   In the Response Phase, each player plays as a Responder. After evaluation of the situation he can stay with his own proposal, or accept a proposal from another player, if it is available. If he accepts a proposal from another player, his own proposition is canceled.
4. Scoring
   The final evaluation is fully automatic; players do not need to act in this phase. The game server automatically computes the best possible sequence of moves for each player and assigns personal scores. Points are calculated as follows. For reaching the goal location a player receives 125 points. If he does not reach the goal, 25 penalty points are subtracted for every square between the goal and the player's position. In addition, for every chip the player has not used, he receives 10 extra points.

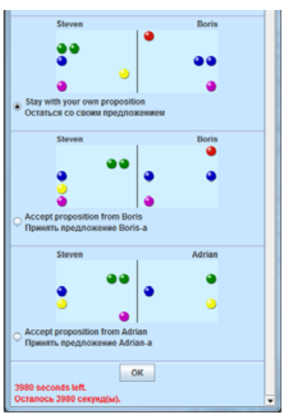


Fig. 4. Screenshot of Colored Trails Game, Response Phase used for 520days isolation

The introduced modifications provide several advantages.

First, we increased number of Proposer per game (from 2 to 3) and, as a consequence, we get more data about the behavior of Proposers.

Second, we increased the number of Responders (from 1 to 2 or 3). This way more information about behavior of Responders can be collected.

Third, we increased the variety of situations in which Responders can be. In particular, Responders have to choose from different number of offers coming from Proposers which are in different situations.

Forth, based on the analysis of the previous experiment, we added a phase at the game, which is aimed to assess irrational preferences of the players. At this stage, when each player is choosing a partner for a current game, player's behavior can not be based on rational thinking about state of the game.

We also designed the Colored Trails Questionnaire that is aimed to define which strategy each player has kept. The players are asked to fill it in after eacj gaming session.

In addition to the previous experiment we combine the usage of the negotiation game with direct monitoring techniques of the nonverbal behavior of the subjects. In that way we want to study if our conclusions about user's persistence of the game and other player can be confirmed by the direct measurement methods. In particular, the monitoring techniques can be used to detect if players are satisfied or not with offers from other players as well as decisions made by the Responders.

3.3 Self Assessment Techniques

Self assessment is an important component of mental health assistant since subjects can explicitly provide needed information about intra- and inter- personal states.

3.3.1 Big 5 Test

One of the intentions of our experiment is to find a relation between observed dynamics of the interpersonal relations with psychological parameters of the crew members. Such a relation could help us to generalize the behavior observed in particular crew and to predict, in this way, relations in groups consisting of members of similar psychological types.

At the baseline of the experiment, we are collecting data, which could give us the psychological characteristics of participants about personality and collaboration patterns at the same time. This data we can be used as a reference point in the future data analysis. For this purpose we are using the Big 5 Test [12].

The Big Five model is considered to be one of the most comprehensive, empirical, data-driven research findings in the history of personality psychology. Over three or four decades of research, these five broad factors were gradually discovered and defined by several independent sets of researchers.

To address the need for a short instrument measuring the prototypical components of the Big Five that are common across investigators, the Big Five Inventory (BFI) has been constructed [10]. The 44-item BFI was developed to represent the prototype definitions developed through expert ratings and subsequent factor analytic verification in observer personality ratings. The goal was to create a brief inventory that would allow efficient and flexible assessment of the five dimensions when there is no need for more differentiated measurement of individual facets.

For our study the results of this test are used to build a Social Network Model, and use it for the further data analysis in terms to compare and find a correlation with the Social Network Model, that we would have from records of the game.

3.3.2 Self Assessment Manikin (SAM) and Cognitive Task Load

To assess the emotional parameters during the game, players will be asked to fill in the questionnaires (see Figure 5).

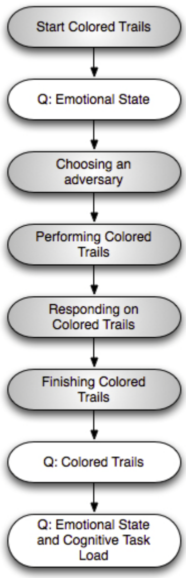


Fig. 5. Colored Trails workflow

For emotional state assessing of the participants, we are using the Self-Assessment Manikin test. The Self Assessment Manikin (SAM) Test was introduced in 1985 by P.J.Lang. It offers the ability to avoid the verbal expression of the emotion assessment, so it establishes a quick and easy to use experimental procedure [13, 14].

Usage of this technique has few advantages. The test is well established, it's aimed for the main and basic parameters, but the results are full enough for analysis. And it's simple and intuitive to perform, which is crucial for our design.

The test assesses the follow emotional states: valence (pleasure), arousal (activity level), dominance (control in the situation).

Cognitive load theory has been designed to provide guidelines intended to assist in the presentation of information in a manner that encourages learner activities that optimize intellectual performance. At the end of each game, the participants are asked to fill out a Cognitive Task Load (CTL) Questionnaire (apart the SAM Questionnaire).

Within the CTL we assess: level of information processing, task-set switches and time occupied.

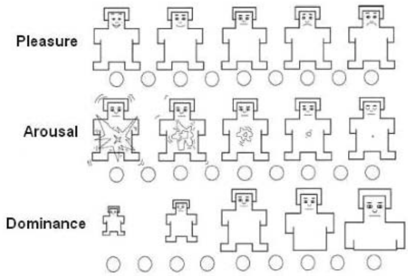


Fig. 6. SAM Questionnaire

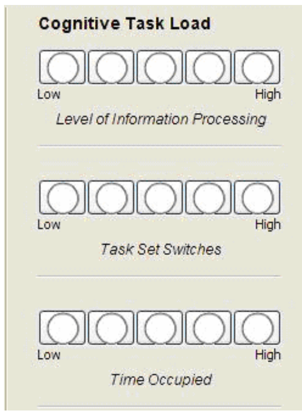


Fig. 7. CTL Questionnaire

Analysis of the SAM before and after each game will show how game could affect an emotional state of the player. Also it would allow us to analyze the correlation between an emotional state of the player and his/her strategies of playing.

3.4 Measuring Nonverbal Behavior

Automatic monitoring of face expression, voice intonation and body movements were proven to be useful techniques for monitoring of emotional states of subjects. These techniques are important since humans naturally express their emotions throw non-verbal behavior.

In our study we use records of facial expressions during the time people are playing the Color Trails Game. Facial expressions give important clues about emotions. In contrast to self assessment techniques, the tracking of nonverbal behavior provides information about emotions and psychological state unconscious. It would allow us to get more objective data compare to results of questionnaires.

Using the software for analysis video (like Visual Recording by University of Amsterdam), we would be able to quantify the 6 main emotions [9].

3.5 Social Network Analysis

Finally, to describe the relation between participants and analyze how the relations are developing through the time, we use Social Network Analysis (SNA). The objective of SNA is to understand the pattern and content of the interactions that take place within and between social units. For this reason, such analysis brings fresh methodological and conceptual power to tests of the contact hypothesis. SNA defines networks as sets of ties linking several individuals. Ties or contacts may be of different kinds, formal or informal, frequent or infrequent, affect-laden or purely utilitarian. Network analysts use the terms "transactional content" or "tie type" to identify the type of exchange or relationship that takes place between actors in a network [15]. SNA approach allow not only to draw a 'map of ties' between structures/ institutions/ societies as macro-level of social reality, but, in addition, to track changes that occur with these 'actors' during social transformations in the contexts of interconnections at the micro-level (meaning connections between individuals, households or organizations).

In this way SNA provides a powerful method for description and representation of results, collected by different kind of measuring techniques, used in the present study.

4 Discussion

4.1 Mars-500 Experiment: 520 Days Isolation

By now, we combine two approaches to measure mental states of the isolated groups of people in attempt to predict arising conflicts in the group. The first is based on analysis of the subjectivity during play of a negotiation game which is verified with the second approach that relies on direct measurements of the nonverbal behavior of the participants. Our initial analysis of the game subjectivity outcomes were based on the 105 days trial. The analysis showed that there were not significant deviations from the logical choices of the proposers or the responders. This may mean that not extreme situations or noticeable changes in the emotional state of the players took place during the 105 days isolation. However, the reason could be that the negotiation game outcomes by itself are not informative about the emerging conflict situations. The highly educated and performance oriented astronauts trainees may be able to suppress their emotional tension during the task of a game play.

We expect that during the longer isolation period the relations between crew members can become more intense. Related experiments have shown that typing intervals, facial expressions and gestures give clearer indication about the emotional tensions, so we can correlate the game outcomes results with these measurements. In connection with the Mission Execution Crew Assistant (MECA) by TNO [19, 20], we have an unique possibility to test our design within MARS- 500 experiment. MARS-500 provides a unique test platform, because of its setting in which a small crew is isolated for a long duration to simulate a manned Mars mission. In this setting, more prolonged or repeated usage of MECA/AMHA can be tested. A major focus of MECA/AMHA is to support team resilience by monitoring team member's performance (i.e., the effectiveness and efficiency of operations and the related refreshment trainings for nominal and off-nominal situations) and corresponding condition (i.e., the appropriateness of his or her cognitive and affective responses). In the MARS-500 program, we will select a small set of core elements of MECA/AMHA that need this type of prolonged evaluation: a diagnostic method to measure psychosocial crew status, a feedback mechanism, and support for crew planning. The objective of this evaluation is to improve the requirements baseline and its design rationale for these elements, and to refine the corresponding models and methods. The setting is as follows:

- Two groups of three astronauts who train and game once a week for half an hour (incl. procedure training and an entertainment game).
- The astronauts communicate via chat.
- MECA/AMHA collects information on crew condition (social network, emotional state) and performance (effectiveness and efficiency of operations during the training and gaming).
- MECA/AMHA provides (simple) feedback on crew condition and performance.

4.2 Web-Based Experiment

To get quantitative data, we are running an additional web-based study. For this purpose the web-site with Colored Trails Game has been developed.

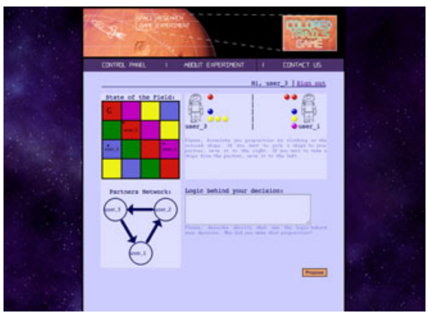


Fig. 8. Screenshot of the Colored Trails Game web site

The research questions for this experiment, apart to get quantitative data, are:

- how the CT game can develop the patterns of existing relationships between people;
- how the CT game can create the relationships between people;
- how the CT game can monitor the behavior.

In these settings we are planning:

- 5-6 groups (by 3 person each) people who know each other
- 5-6 groups (by 3 person each) people who don't know each other
- About 15-20 groups of three people, which are know each other well and/or working together.
- Three sessions per week.
- The duration of the experiment is 15-18 sessions (5-6 weeks).

By the end of the experimental part of the project, we would be able to do data analysis. Two different experimental settings provide an opportunity for cross validation and answer the question of transferability of results obtained with a single experiment with fixed constrains.

Acknowledgments. The AMHA project is supported by NWO User Support Program Space Research. The project number is ALW-GO-MG/07-13.

The authors would like to thank the management and the work group of the Mission Execution Crew Assistant (MECA) project, and particularly Prof. Mark Neerincx, for support and collaboration.

We also are grateful to Vadim Guschin, the chief researcher at the Institute of Biomedical Problems, for support with preparation and conduct AMHA-experiment for Mars-500 study.

References

1. Kanas, N., Salnitsky, V., Grund, E., Gushin, V.: Social and cultural issues during Shuttle/Mir space missions. Acta Astronautica 47(2-9), 647–655 (2000)
2. Manzey, D.: Human missions to Mars: new psychological challenges and research issues. Acta Astronautica 55, 781–890 (2004)
3. Gushin, V.: Psychological countermeasures during space missions: russian experiences. Journal of Gravitational Physiology 9(1), 311–312 (2002)
4. Hennes, D., Tuyls, K., Neerincx, M.A., Rauterberg, G.W.M.: Micro-Scale Social Network Analysis for Ultra-Long Space Flights. In: IJCAI-09/SMC-IT-09/IWPSS-09 Workshop on Artificial Intelligence in Space (2009)
5. Hennes, D., Tuyls, K., Rauterberg, G.W.M.: State-Coupled Replicator Dynamics. In: 8th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2009), pp. 789–796 (2009)
6. Hennes, D., Tuyls, K., Rauterberg, G.W.M.: Formalizing Multi-state Learning Dynamics. In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (IAT 2008), vol. 2, pp. 266–272 (2008)
7. Rauterberg, G.W.M., Neerincx, M.A., Tuyls, K., Loop, J.: Entertainment computing in the Orbit. New Frontiers for Entertainment Computing 279, 59–70 (2008)
8. Gaal, Y., Grosz, B.J.: Colored Trails: A Formalism for Investigating Decision-making in Strategic Environments (2005)
9. Valenti, R., Sebe, N., Gever, T.: Facial Expression Recognition: a Fully integrated approach (2007)
10. John, O.P., Srivastava, S.: The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives (2005)
11. Camerer, C.F.: Behavioral Game Theory: Thinking, Learning, and Teaching (2001)
12. Harvey, R.J., Murry, W.D., Markham, S.E.: A ”Big Five” Scoring System for the Myers-Briggs Type Indicator (1995)
13. Bradley, M.M., Lang, P.J.: Measuring Emotions: the Self-Assesment Manikin and the Semantic Differential. J. Behav. Ther. and Exp. Psychiat. 25(I), 49–59 (1994)
14. Tsonos, D., Kouroupetroglou, G.: A Methodology for the Extraction of Reader’s Emotional State Triggered from Text Typography. Tools in Artificial Intelligence, 139–154 (2008)
15. Nelson, E.: The Strength of Strong Ties: Social Networks and Intergroup Conflict in Organizations. The Academy of Management Journal 32(2), 377–401 (1989)

16. Busso, C., Deng*, Z., Yildirim, S., Bulut, M., Lee, C.M., Kazemzadeh, A., Lee, S., Neumann*, U., Narayanan, S.: Analysis of Emotion Recognition using Facial Expressions. Speech and Multimodal Information (2004)
17. Mase, K.: Recognition of facial expression from optical flow. IEICE Transc., E. 74(10), 3474–3483 (1991)
18. Yacoob, Y., Davis, L.: Computing spatio-temporal representations of human faces. In: Proceedings of IEEE Computer Society Conference, Computer Vision and Pattern Recognition CVPR 1994, pp. 70–75 (1994)
19. Neerincx, M.A., Lindenberg, J., Smets, N.: The Mission Execution Crew Assistant: Improving Human-Machine Team Resilience for Long Duration Missions (1998)
20. Breebaart, L., Bos, A., Grant, T., Soler, A.O., Neerincx, M.A., Lindenberg, J., Smets, N., Brauer, U., Wolff, M.: The MECA Project — Ontology-based Data Portability for Space Mission Operations (2008)
21. Barakova, E., Lourens, T.: Expressing and interpreting emotional movements in social games with robots. In: Personal and Ubiquitous Computing (2010)
22. Lourens, T., Barakova, E.: Retrieving Emotion from Motion Analysis: In a Real Time Parallel Framework for Robots. In: Leung, C.S., Lee, M., Chan, J.H. (eds.) ICONIP 2009, Part II. LNCS, vol. 5864, pp. 430–438. Springer, Heidelberg (2009)
23. Lourens, T., Barakova, E.: My Sparring Partner is a Humanoid Robot -A parallel framework for improving social skills by imitation. In: Mira, J., Ferrández, J.M., Álvarez, J.R., de la Paz, F., Toledo, F.J. (eds.) IWINAC 2009. LNCS, vol. 5602, pp. 344–352. Springer, Heidelberg (2009)
24. Gal, Y., Grosz, B.J., Kraus, S., Pfeffer, A., Shieber, S.: Colored trails: a formalism for investigating decision-making in strategic environments. In: Proc. of the IJCAI Workshop on Reasoning, Representation, and Learning in Computer Games (2005)
25. Ficici, S., Pfeffer, A.: Modeling how humans reason about others with partial information (2008)

Engaging Autistic Children in Imitation and Turn-Taking Games with Multiagent System of Interactive Lighting Blocks

Jeroen C. J. Brok and Emilia I. Barakova

Eindhoven University of Technology
ID, Den Dolech 2, POB 513,
e.i.barakova@tue.nl

Abstract. In this paper game scenarios that aim to establish elements of cooperative play such as imitation and turn taking between children with autism and a caregiver are investigated. Multiagent system of interactive blocks is used to facilitate the games. The training elements include verbal description followed by imitation of video-modeled play episodes. By combining this method with the tangible multiagent platform of interactive blocks (i-blocks) children with autism could imitate play episodes that involved turn taking with a caregiver. The experiment showed that most of the children managed to imitate the play scenarios after video modeling, and repeat the behaviors with the tangible and appealing block platform. When all the actions were well understood by the autistic children, they performed willingly turn taking cooperative behaviors, which they normally do not do.

Keywords: Games, collaborative play, tangibles, i-blocks, autism, behavioral training, turn taking, imitation.

1 Introduction

Teaching of socially relevant behavior to children through games with robotic toys is an area of emerging interest [3, 4, 6, 9, 21, 25]. By robotic toys we understand autonomous robots and tangibles that are equipped with sensors and actuators, mediated by a computational device. They can perceive and act upon the surrounding environment that may include other robotic toys or humans and in this way they have the affordance to be used as a mediator of play and education [3, 4, 6, 9, 21, 24, 25, 26].

A number of robotic toys, both with and without anthropomorphic or zoomorphic shape have been used in games with autistic children [5, 7, 9, 20, 21, 22, 25, 26]. The autonomous robotic toys in these studies, perform simple behavior (as for instance lifting of a hand) that aims to provoke reciprocal human reaction [9, 10, 21], to predict the position or the direction of the movement which is referred to as predicting motor intentions [5, 9, 21, 29], and become a mediator of play [3, 6, 26].

Robotic toys have been shown to engage children with autism at various age levels [3, 16, 25, 27] because of their predictable and controlled behavior, their physical embodiment and because of the appeal that computerized games have to people with autism. Tangibles, in particular, have been shown to support collaboration by typical

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 115–126, 2010.

children [8, 17]. Experiments with "Lego Therapy" [19, 23] suggest that interaction with physical building blocks may support autistic children to collaborate for extended periods of time. In discussions with caregivers in different autistic institutions we found out that it is a common practice to turn abstract concepts like time and mood to visible or tangible categories, by using for instance adapted clocks for visualization of the time and traffic light to express moods through colors.

We use a multiagent platform of interactive blocks (i-blocks) that express behavior in accordance with the way the blocks are used [2, 3, 4]. The platform has a resemblance with tangible interaction platform, because of its appearance, and with a multi-agent system of robots, since it expresses behavior through changing the lighting pattern of the blocks in emergent manner [4]. It has shown to be very engaging for children with autism, because of the following reasons. Blocks are a usual toy for children with autism. The regular cubic shape enables the children to form a grid in order to make the spatial relations more obvious. In addition, the cubes can be stacked, enabling more complex structures. The blocks that we have developed emit dimmed colored lights which change in accordance with the way the blocks are used. Light stimulation is by itself known to be appealing to the autistic children. In addition, the puzzling lighting behavior motivates the autistic children to find out the underlying rule of interaction of the blocks and to explore with the platform in order to discover new emergent behaviors.

Nadel [21] showed, that imitation, turn taking, and recognition of mutual action patterns is a basis for developing the sociall skills and inferring others' intention correctly. Inspired by these findings, in the current study we use the i-blocks platform for turn taking and imitation games.

Earlier research indicated that if the children with autism are asked to make a verbal description of the acted out target behavior before imitating it helped them to sustain cooperative play, which resulted in longer play episodes and more variation in play [18]. It was shown that explaining beforehand the performed behavior to the autistic child makes possible more complex imitation to be performed. Different studies have established that autistic children would imitate either the goal of the imitation, or the particullar movement behavior. Combining both, however have shown to be challenging for the children with autism.

Therefore we prepared scenarios that contained relatively complex imitation behaviors in which the children had to understand, verbalize, and imitate these behaviors. Similar to Jahr and colleagues [18], we used a person to explain the behavior that has to be imitated to the children. Inspired by studies of videomodelling, such as the ones reported in [11, 12, 13, 14, 28], instead of using actors or the same individual to show the behavioral scenarios every time, we made videos of the demonstrated scenarios.

We developed cooperative play scenarios that involved imitation and turn taking between children with autism and a caregiver. In our approach to training children with autism, we target the possible complex of reasons that may cause the atypical or missing cooperative play. First, we address the lack in variation during play that makes the child's development sometimes one-sided and limits social interaction with peers. Second, we also consider the difficulties of autistic children to stay concentrated and keep their attention to the current activity. Third, we focus on imitation of the targeted behavior, instead of the individual movements.

2 Training Social Play Behaviors by Autistic Children.

In an investigation of social play behavior Parten [24] distinguishes between several types of play, depending on the level of social involvement of the children, namely *solitary independent play, parallel activity, associative play, and cooperative or organized supplementary play.* These play behaviors were found typical for different age groups.

In *solitary independent play* and *parallel activity* the child plays independently from other children. By parallel activity the child is situated among other children and plays with toys that are like those which the other children are using, but does not try to influence/modify the activity of the nearby children. In *associative* and *collaborative play* the child plays with other children. There is a greater level of awareness of the peers and there is a borrowing and loaning of play material; there is evident interaction, as for instance, following one another with trains or wagons. By collaborative play there is clear exchange of play objects.

Autistic children are most often observed to be engaged in *solitary independent play* and some in *parallel activity,* even when their age progresses. We aim to train elements of *associative* or *collaborative play* through appropriate games. Problems in the social aspects of play in autism can be traced to the earliest forms of social interaction with the caregivers and may include a range of behaviors. We consider imitation behavior as central to social learning and development.

Imitation does not occur naturally in social development of people with autism. Therefore, imitation behaviors have been used in many different ways to help autistic children improve their social skills. For instance, children with autism do not imitate naturally but like being imitated. A typical experimental scenario requires a caregiver to imitate the autistic child in order to engage him/her in social interaction [15, 17]. Aldridge and coleagues [1] conducted experiments with autistic people that aim to differentiate the elements of imitation that are not deficient by autistic people. They constructed an experiment with 'unfulfilled intentions' tasks and found, much to their surprise, that the autistic children (mean age = 39 months) did imitate the goal of an action very well. Another studies showed that although the children can imitate the goal of an action, they would either fail on mimicry, i.e. repeating the actions, if the relation between those actions and the outcome they produce is not clear, or they fail by imitation proper, i.e. the imitation of both: the means and the goal of an action.

We designed an experiment where the goal of the imitation has to be achieved through performing a sequence of simple turning movement behaviors.

Next to imitation, facilitating turn taking is particularly important to the children with autism because of their weak social responsiveness. Sensitivity to the contingencies involved in cooperative play may produce turn taking and variability in the partners play behavior.

To address the problem of turn taking through the imitation of the goal of an action, we used a task similar to those of Jahr [18] where actors were demonstrating actions to the children. To minimize the stress levels of newnes due to meeting actors that can behave differently every time we asked a single actor to demonstrate all the scenarios. Both, the request for social interaction and the newness of the actor can make it unnecesary stressfull for the child to attend the training. To further resuce this stress, we created video scenarios instead of using physically present actor, similar to Charlop-Christ and colleagues [11] .

In summary, our approach is expected to have the following advantages. (1) The same actor can do every scenario, which allows the child to develop a bound with the actor. (2) The child won't feel as stressed watching a video as he or she will be while watching live acted scenarios. There is no social response expected and therefore the child will be more relaxed. (3) The tape can be distributed anywhere. (4) One can focus more on the important aspects in the video, by zooming in at highlights or blurring out unwanted stimuli. This will keep the child from getting distracted by unimportant details. (5) Video modeling will also allow the actors to shoot an optimal scenario. (6) There will be less people required to provide the necessary support for the children.

3 Experimental Design

3.1 Experimental Procedure

The experimental procedure contains out of three steps that are repeated every time. (1) The autistic child watches the video of the target behavior together with a coordinator. The coordinator will make sure the child pays attention while the actors in the video play out the scenario step by step. (2) The coordinator lets the child describe what was shown in the video scenario. This procedure has to be repeated until the child describes the whole scenario without errors. (3) The coordinator will than imitate the scenario together with the child. After completion of the scenario without errors the video with increased complexity (for instance increased number of play responses) can be presented.

Teaching autistic children to imitate at a very early age can be crucial for their development. Therefore we target very young children. As shown by [18] two important requirements considering the children have to be satisfied. (1) The child has to be able to follow instructions and perform play responses with toys upon verbal instructions. (2) The child has to be able to describe play activities. After watching the video the child has to be able to describe what he or she has seen.

The quality of the video might also influence the results. Earlier research [17] has shown that it is not important whether the actors are adults or children. However, the quality of the act is essential. The actors have to be trained to perform a list of scripted play episodes. This list has to be assessed by an expert to define whether this is an appropriate play scenario. The first video introduced the blocks, the 3 interaction possibilities, and a play scenario for the child to practice individually. The second video was a scenario with two play responses; the child imitated one action from the scenario than shared his play area and let the other person complete the scenario. The final video showed four play responses which required an additional turn taking behavior than the second video.

3.2 The Multiagent Platform of Interactive Blocks –i-Blocks

We use multiagent system of i-blocks [2, 3, 4], which interact when positioned in each other's vicinity (Figure 1). They emit colored light and depending on the algorithm that

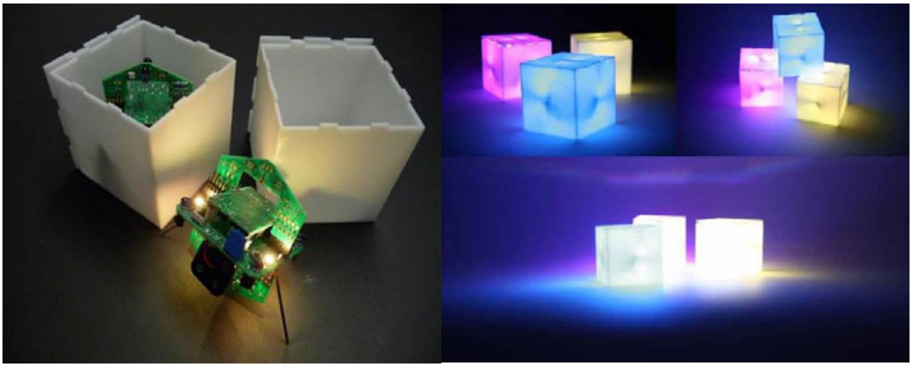


Fig. 1. The developed multi-agent system of interactive blocks (a) internals of the block platform and (b) example of emergent light patterns when the blocks are put in each other's vicinity

is loaded on each block at this moment, express different set of local interaction behaviors that cause emergent collective behaviors. The overall behavior of the system depends on the local interactions, and therefore it forms an embodied multiagent system [4]. The complexity of the emergent behaviors depends on the complexity of the individual behaviors of the blocks.

We intended to create games that are challenging, but still not too unpredictable that the child wouldn't understand the behavior. Cooperative play can be assessed by the number of play responses. By encouraging turn taking and motivating the players to work towards achieving a common goal, we assume that the play responses will increase.

There was one active block (which can change the colors of other blocks) and 5 passive blocks, that can be subject to a color change. There are sheets of colored paper at the 4 corners of the table. The task for the children was to match the color of the blocks to the pieces of colored paper positioned in front of the children. For example a colorless passive receiving block laying on red piece of paper should be colored red. This can be done by turning the active sending block to red and holding it close to the passive receiving block. Turning the active block causes a change of its color and this is the behavior that the child has to imitate.

Examples of different behavioral possibilities of the blocks in the framework of different game scenarios are described in [2, 3, 4]. In [2] the technical features of the platform are described. The blocks can interact by detecting their neighbors through the built-in sensors, and show behavior in response to their sensory stimulation. A build-in accelerometer is used to record the children's hand movement and to facilitate the imitation game needed for the current study.

The blocks were specially designed to fit the play habits and the patterns of thinking of the autistic children. Initial user tests as reported in [3, 4] have shown that children find them very engaging and pleasurable.

3.3 Play Scenarios

Three scenarios has been performed and tested in the following way. First, the child and the coordinator view the video scenario on a computer or TV (TV offers less distractions; computer keyboard and mouse distract the child). After that the coordinator asks the child to describe what actions were performed by the actor. If the scenario is described correctly the child can start imitating the scenario, if not the video scenario is re-played until the child explains correctly the actions. The child plays the scenario with the coordinator.

Scenario one: The blocks are placed in front of the actor, the passive blocks are on the colored pieces of paper, and the active block is in the middle. The actor picks up the active block and starts turning it until the block gets the color of the paper in front of him. The actor transfers the color to the passive block on the piece of paper. He does so for all the three paper-block combinations. The scenario is finished when all passive blocks are colored according to the underlying piece of paper.

Goal: The child gets familiar with the interaction possibilities of the blocks and the concept of the game.

Implementation: The blocks are now placed in front of the child. The child has to perform the following actions:

1.1 Child watches video
1.2 Child describes scenario
1.3 Child colors first block
1.4 Child colors second block
1.5 Child colors third block

The game is finished when all passive blocks are colored according to the underlying piece of paper.

Scenario two: In this scenario the actions were performed by two actors. The blocks are placed between the two actors. The first actor picks up the active block and starts turning it until the block becomes the right color. He transfers the color to the passive block on the piece of paper. The first actor shares the block with the second actor who, on his turn, colors his block.

Goal: let the child complete the most basic cooperative scenario, where the child shares the block once with the coordinator.

Implementation: After the video the coordinator asks the child what actions were performed by the two actors, and how they responded to each other. The blocks are now placed between the child and the coordinator. The following actions take place:

2.1 Child watches video
2.2 Child describes scenario
2.3 Child colors first block
2.4 Child shares block (cooperates)
2.5 Coordinator colors a block

The game is finished when the child shares the block with the coordinator who colors his or her block in the same way.

Scenario three: This is an extension of *Scenario two* where the active block is shared twice.

Goal: let the child complete a longer scenario, where he/she shares the block two times and receives the block once.

Implementation:

3.1 Child watches video
3.2 Child describes scenario
2.3 Child colors first block
3.4 Child shares block (cooperates)
3.5 Coordinator colors block
3.6 Coordinator shares
3.7 Child colors 2nd block
3.8 Child shares block (cooperates)
3.9 Coordinator colors 2nd block
The scenario ends after the coordinator colored his or her second block.

Fig. 2. Snapshots of performing scenario 3. a) Child shares block (cooperates) b) Coordinator colors 2nd block after the second turn taking.

4 Results

4.1 User Observations

To verify the experimental choices, we conducted a confrontation of autistic children with the platform and the experimental conditions. All the participants were aged between six and eight. They were diagnosed with some form of autism. The first two participants didn't see the introducing video, but were allowed to explore the behavior of the blocks together just to find out whether the introduction video was necessary. The children enjoyed the free playing, but didn't manage to perform scenarios 2 and 3. Participant 1 (P1) found experimenting far more interesting than watching the videos. He was constantly distracted from the video by playing with the blocks. Although P1 got distracted several times during the video he and P2 were able to complete the scenarios. The rest of the children played together with the coordinator instead of playing with peers.

P3 had no difficulties understanding the games and even figured out the final game before the video was completely finished. P4 was very focused while watching the video scenarios. After fully describing the games P4 imitated them perfectly without any mistakes. P5 was very involved in the videos; he counted the blocks together with the actors in the videos and named the colors like the actors did. He was not only very involved, but also playful and took initiative to explore. After completing the scenarios without any problems P5 started to build towers and explore the blocks interactive possibilities. P6's comments and repetitive behavior gave the impression that she tries to remember colors by establishing a pattern. P6 was the only one who needed to re-watch the scenarios about turn taking to complete them without mistakes.

The scenarios are in general too easy for children of this age group and therefore should be either tested with younger children or made more difficult and challenging for this group. As seen in Table 1 the blocks brought a high level of distraction for the children while they were watching the scenarios.

Table 1. The number of distractions per participant

Participant	Total distractions	Distracted by blocks
P1	13	13
P2	2	0
P3	6	1
P4	4	3
P5	12	9
P6	10	[illegible]

Another important finding was that it is hard to make the children exactly describe the actions. The child possibly has to be informed before the video starts that he or she will have to describe the actions in the video. In the current setting the children often already started playing during the video or directly after it.

4.2 User Test

The user test was performed with five autistic children of 4 and 5 years of age. Two of them were diagnosed with PDD-NOS (P1,P2), one PDD-NOS or classic autism (P3), one classic autism and ADHD (P4) and for the fifth child there were no definite outcomes of the diagnosis yet (P5). All the children managed to finish two or three of the scenarios.

The children were prepared for the test by their development coordinator, who also guided the children from their classroom to the test room and imitated the scenarios with the children. At the end she wrote a qualitative evaluation of the tests which will be summarized below.

P1 didn't have much trouble with the games. He finished all three scenarios without any errors. There was only one short distraction when the child looked up from the third video to check out the blocks. The rest of the videos were watched without any distractions and without any necessary encouragement from the coordinator.

P2 was very enthusiastic about playing with the blocks, despite he normally does not show variation in play. He wanted to start playing right away, but after calming down he watched the videos without more than one distraction moment looking over at the blocks. An interesting observation was that P2 who normally has trouble sustaining one on one play, was now concentrated on the scenarios and followed the instructions. During scenario three he first didn't shared the block with the coordinator, but eventually after asking what he had seen in the video he completed it by involving the coordinator in the play scenario.

P3 usually doesn't show initiatives in play, he is very passive and has to be stimulated a lot. Watching videos was one of his favorite activities and therefore he became enthusiastic for our scenarios very quickly. His approach during the session was more active than normally in the group, he counted together with the actress in the video and mentioned the colors like she did. The final scenario was completed after watching the video second time because the first time the imitation that P3 showed was not correct.

P4 normally shows minimal group play and play variability – he plays every day with one and the same train. It took more effort to get him to participate in the session. The child had to be stimulated several times to keep watching the video. It was surprising that he followed the instructions and let the coordinator join in the game. The last most difficult scenario was too hard for him. After watching the video for the second time the child didn't manage to complete the scenario, because he became heavily distracted.

P5 already shows quite varied play, but still had turn taking problems. He was very enthused and explorative with the blocks. He was focused on the videos and showed turn taking during all the scenarios without needing further instructions or visualization.

In Table 2 the number of distractions in the test are shown. Although the total distraction still varies widely per person the distractions by the blocks considerably decreased because of the better preparation of the test. The total number of distractions also went down due to better preparation of the coordinator.

Table 2. The number of distractions per participant.- improved experimental design

Participant	Total distractions	Distracted by blocks
P1	7	2
P2	1	0
P3	3	0
P4	21	3
P5	1	0

5 Discussion

Despite the lack of quantitative results the following observations suggest that the proposed method shows a potential in supporting autistic children in learning imitation and turn taking behaviors at a very early age. (1) Most of the children managed to imitate the play scenarios. (2) The videos showed to be a suitable way to teach the children understand and imitate the target behavior. (3) The stress levels of the children stayed lower than in actual social contact with new person, as observed by the coordinator and children could get prepared well for the upcoming scenario.

Earlier studies have shown that the basic principles of this method, can improve the social skills of the children. We changed the setting by introducing multiagent system of tangible interaction blocks that were especially designed for autistic children and were tested to be perceived as very pleasurable and engaging. It is common for children with autism to choose the play objects based on the sensory stimulation that they provide, such as color, touch, sound, or smell. The i-blocks, which emit pleasant dimmed light, seen through the semi-transparent walls, has been well accepted by all autistic children that has been participating in the experiments so far. Implementing an accelerometer within the i-blocks makes them sensitive to handling with a hand and makes possible to implement and test imitation and other interaction behaviors that are rooted in motor handling actions. We are able to test in the future the ability of the children of mimicry or blind imitation, of goal imitation or of the overall process of imitation proper. At this study goal imitation was used to further facilitate turn taking. Both imitation and turn taking are components of cooperative play behaviors. We showed that when all the actions are well understood by the autistic children, they performed willingly turn taking behaviors, which they normally do not do.

We plan a longitudal study with an extended range of imitation and turn taking games that will aim at testing whether the children will be able to transfer the learned cooperative play behaviors to different, preferably real life situations. Improvements in quality and quantity of the videos will be necessary to keep the children interested on longer term. Within the current set-up the video is acted out by two amateur actors. For a final result the video scenarios need professional actors with prior experience in making children videos. These actors will be required to make the video perfectly understandable and enjoyable for the children. The set up should also be more professional regarding the lighting conditions of the room where the video material was taken and professional editing can also remove some minor flaws from the videos.

In the current research the play of a child and a caregiver in a prepared environment has been observed. We would like to explore the effect of this method in natural environment and in realistic every day interactions, such as at school with autistic and typical children.

Acknowledgments

We would like to thank Dr. Juliane Cuperus, Health psychologists/ neuropsychologist from Sint Marie hospital in Eindhoven, the Netherlands, for her evaluation collaboration. We also thank Rachel Delhees and Marleen Vissers for their collaboration setting up and carrying out the tests. Finally we thank Het Palet school in Weert, the Netherlands for their support.

References

[1] Aldridge, M.A., Stone, K.R., Sweeney, M.H., Bower, T.G.R.: Preverbal children with autism understand the intentions of others. Developmental Science 3, 294–301 (2000)

[2] Alers, S.H.M., Barakova, E.I.: Multi-agent platform for development of educa-tional games for children with autism. In: IEEE ICE CIG 2009, pp. 47–53 (2009) ISBN: 978-1-4244-4459-5

[3] Barakova, E.I., Gillessen, J., Feijs, L.: Social training of autistic children with interactive intelligent agents. Journal of Integrative Neuroscience 8(1), 23–34 (2009)

[4] Barakova, E.I., van Wanrooij, G., van Limpt, R., Menting, M.: Using an emergent system concept in designing interactive games for autistic children. In: Proc. of IDC 2007, Aalborg Denmark, June 2007, pp. 73–77. ACM, New York (2007) 978-1-59593-747-6

[5] Barakova, E.I., Vanderelst, D.: From spreading of behavior to dyadic interaction—a robot learns what to imitate. Int. J. Intell. Syst. (2020) (in press)

[6] Barakova, E.I., Lourens, T.: Expressing and interpreting emotional movements in social games with robots. In: Personal and Ubiquitous Computing, pp. 1–11. Springer, Heidelberg (2010)

[7] Bekker, M.M., Sturm, J., Barakova, E.I.: Designing for social interaction through physical play. In: Sturm, J.A., Bekker, M.M. (eds.) Proc. Design for social interaction through physical play, Eindhoven, The Netherlands, October 22 2008, pp. 7–10 (2008)

[8] Bekker, T., Sturm, J., Eggen, J.: Designing Playful Interactions for social interaction and physical play. Personal and Ubiquitous Computing (2010), doi:10.1007/s00779-009-0264-1

[9] Billard, A., Mataric, M.J.: Learning human arm movements by imitation: Evaluation of a biologically inspired connectionist architecture. Robotics and Autonomous Systems 37(2-3), 145–160 (2001)

[10] Breazeal, C., Buchsbaum, D., Gray, J., Gatenby, D., Blumberg, B.: Learning from and about others: towards using imitation to bootstrap the social understanding of others by robots. Artif. Life 11(1/2), 3–62 (2006)

[11] Charlop-Christ, M.H., Le, L., Freeman, K.A.: A Comparison of Video Modeling with In Vivo Modeling for Teaching Children with Autism. Journal of Autism and Developmental Disorders, 537–552 (2004)

[12] Corbett, B.A.: Video Modeling: A Window into the World of Autism. The Behavior Analyst Today 4(3) (2003)

[13] Corbett, B.A., Abdullah, M.: Video Modeling: Why Does It Work for Children with Autism? Journal of Early and Intensive Behavior Intervention 2(1), 2–8 (2005)

[14] D'Ateno, P., Mangiapanello, K., Taylor, B.A.: Using Video Modeling to Teach Complex Play Sequences to a Preschooler with Autism. Journal of Positive Behavior Interventions 5(1), 5–11 (2003)

[15] Escalona, A., Field, T., Nadel, J., Lundy, B.: Brief Report: Imitation Effects on Children with Autism. Journal of Autism and Developmental Disorders, 141–144 (2004)

[16] Feil-Seifer, D.J., Mataric, M.: Robot-Assisted Therapy for Children with Autism Spectrum Disorders. In: Proc. IDC 2008: Children with Special Needs, Chicago, USA (2008)

[17] Field, T., Field, T., Sanders, C., Nadel, J.: Children with autism display more social behaviors after repeated imitation sessions. Autism 5(3), 317–323 (2001)

[18] Jahr, E., Eldevik, S., Eikeseth, S.: Teaching children with autism to initiate and sustain cooperative play. Elsevier Science Inc., Amsterdam (2000)

[19] LeGoff, D.: Use of LEGO as a therapeutic medium. Journal of Autism and Developmental Disorders 34(5), 557–571 (2004)

[20] Lourens, T., Barakova, E.I.: My sparring partner is a humanoid robot. In: Mira, J., Ferrández, J.M., Álvarez, J.R., de la Paz, F., Toledo, F.J. (eds.) IWINAC 2009. LNCS, vol. 5602, pp. 344–352. Springer, Heidelberg (2009)

[21] Nadel, J.: Early imitation and a sense of agency. In: Proceedings of the 4th international workshop on epigenetic robots (2004)

[22] Nadel, J., Simon, M., Canet, P., Soussignan, R., Blancard, P., Canamero, L., Gaussier, P.: Human responses to an expressive robot. In: Proceedings of Epigenetic robotics 2006. Lund University Cognitive Studies, vol. 128, pp. 79–86 (2006)

[23] Owens, G., Granader, Y., Humphrey, A., Baron-Cohen, S.: LEGO therapy and the social use of language programme. Journal of autism and developmental disorders 38(10), 1944–1957 (2008)
[24] Parten, M.B.: Social participation among pre-school children. Journal of Abnormal & Social Psychology 27(3) (1932)
[25] Robins, B., Dautenhahn, K., Dickerson, P.: Robots as embodied beings—interactionally sensitive body movements in interactions among autistic children and a robot. In: Proc. of ROMAN 2005, pp. 54–59 (2005)
[26] Robins, B., Dickerson, P., Stribling, P., Dautenhahn, K.: Robot-Mediated Joint Attention in Children with Autism: A Case Study in Robot-Human Interaction. Interaction Studies 5(2), 161–198 (2004)
[27] Stanton, C.M., Kahn Jr., P.H., Severson, R.L., Ruckert, J.H., Gill, B.T.: Robotic Animals Might Aid in the Social Development of Children with Autism. In: Procs. ACM/IEEE Int. Conf. on Human-Robot Interaction, Amsterdam, Netherlands (2008)
[28] Thelen, M.H., Fry, R.A., Fethenbach, P.A., Frautsch, N.M.: Therapeutic videotape and film modeling: A review. Psychological Bulletin 86, 701–720 (1979)
[29] Want, S.C., Harris, P.L.: How do children ape? Applying concepts from the study of non-human primates to the developmental study of 'imitation' in children. Developmental Science 5(1), 1–14 (2009)

Multiple Page Recognition and Tracking for Augmented Books

Kyusung Cho, Jaesang Yoo, Jinki Jung, and Hyun S. Yang

Department of Computer Science, Korea Advanced Institute of Science and Technology, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea
{qtboy,jsyoo,jk,hsyang}@paradise.kaist.ac.kr

Abstract. An augmented book is an application that augments virtual 3D objects to a real book via AR technology. For augmented books, some markerless methods have been proposed so far. However, they can only recognize one page at a time. This leads to restrictions on the utilization of augmented books. In this paper, we present a novel markerless tracking method capable of recognizing and tracking multiple pages in real-time. The proposed method builds on our previous work using the generic randomized forest (GRF). The previous work finds out one page in the entire image using the GRF, whereas the proposed method detects multiple pages by dividing an image into subregions, applying the GRF to each subregion and discovering spatial locality from the GRF results.

Keywords: Augmented Books, Multiple Page Tracking, Markerless Visual Tracking, Augmented Reality.

1 Introduction

Recently, there have been a variety of approaches to enhance books by adding digital information. As an example of these approaches, some applications have enhanced real books by means of augmentation with 3D virtual objects using augmented reality technology. We refer to these applications as the augmented books.

Like other augmented reality systems, the most important problem of the augmented book is the registration between the real and virtual worlds. To address the registration problem, most of the augmented book applications employ fiducial markers because they are easy to recognize and track [1],[2],[3]. However, fiducial markers can lead to visual discomfort due to their distinct shapes, so there is a growing tendency to employ markerless page tracking methods [4], [5], [6], [7]. Especially, Cho et al.'s work [7] has presented the markerless tracking method capable of handling a lot of pages (about 200 pages) using the generic randomized forest (GRF).

All proposed markerless tracking methods for augmented books have not so far recognized multiple pages simultaneously. This imposes restrictions on book designs and users' activitie. If multiple sheets of paper can be recognized, more various interactions on an augmented book may be possible. For multiple page recognition and tracking, we propose the novel markerless tracking method which extends our previous work [7]. The previous work finds out one page in the entire image using the

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 127–138, 2010.

GRF, whereas the proposed method detects multiple pages by dividing an image into subregions, applying the GRF to each subregion and discovering spatial locality from the GRF results.

The remainder of this paper is organized as follows. In Section 2, we review the GRF in the previous work. Section 3 focuses on multiple page recognition and tracking of the proposed method. Section 4 presents the experiment results and Section 5 concludes this paper.

2 Generic Randomized Forest

Given an input image by a camera, an augmented book application needs to recognize a page and perform wide-baseline keypoint matching for calculating its initial pose. For this purpose, in our previous work [7], the generic randomized forest (GRF) which extends the original randomized forest (RF) proposed in [8] is presented. An original RF is used to match keypoints against already-trained keypoints coming from one object. It consists of several randomized trees where internal nodes test the intensity difference between two pixels of an image patch around a keypoint and leaf nodes store the probability distribution related to all keypoints.

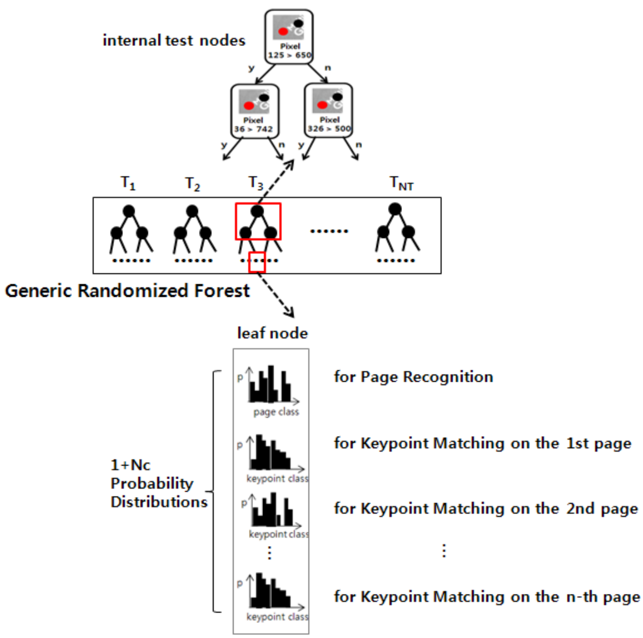


Fig. 1. Generic Randomized Forest

To handle a lot of pages, the GRF stores the probability distribution for page recognition as well as the probability distribution for keypoint matching of each page in leaf nodes like in Fig. 1. If a book has *Nc* pages, each leaf node of the GRF includes *Nc+1* probability distributions. In real-time, N keypoints are extracted from an image by any detector and then passed through the GRF to recognize a current page, as in (1).

$$Page\ \hat{\imath} = argmax_i\, P(C = i|T_1, \cdots, T_{NT}, m_1, \cdots, m_N) \quad (1)$$
$$= \underset{i}{argmax} \frac{1}{N} \sum_{j=1}^{N} \frac{1}{NT} \sum_{t=1}^{NT} P(C = i|leaf(T_t, m_j)) \ ,$$

where $leaf(T_t,m_j)$ is the leaf node which the i-th keypoint m_i reaches in the t-th tree T_t. If the i-th page is recognized as a result of (1), then keypoint matching for the i-th page considers only the i-th probability distribution stored in the leaf nodes. Keypoint m_j is matched, as in (2).

$$Keypoint\ \hat{k} = P\left(K = k\middle|T_1, \cdots, T_{NT}, m_j\right) \quad (2)$$
$$= \underset{k}{argmax} \frac{1}{NT} \sum_{t=1}^{NT} P(K = k|leaf(T_t, m_j))$$

3 Proposed Method

In the following we describe our proposed method capable of recognizing and tracking multiple pages, which is intended for obtaining page information including IDs and pose information of all pages visible in an image captured by a camera. The final page information (**P**) is a set of visible pages' information $p = (ID, R, T)$, where R is a 3x3 rotation matrix and T is a 3-translation vector representing a page pose.

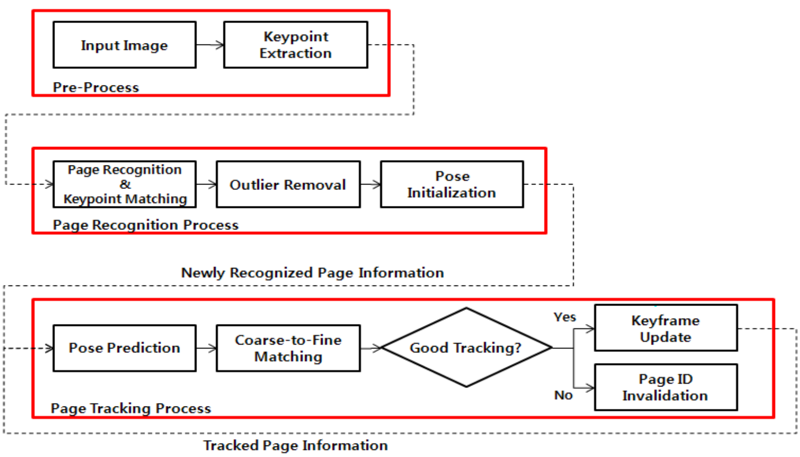


Fig. 2. Overview of the proposed method

Prior to the real-time processing, an offline training process is required. In the training process, we take one picture of each page included in the book, extract the keypoints, and train the GRF in such a way as to be explained in [7].

The proposed method consists of three main processes: pre-process, page recognition process, and page tracking process as shown in Fig. 2. In real-time, keypoints are first extracted from the input image by the FAST detector [9] because it is well known to be very fast. However, keypoints extracted by the FAST detector do not contain scale information, so we build three levels of an image pyramid and extract keypoints from each level. They are conveyed to the page recognition process to recognize pages appearing newly in the image except already-tracked pages. For the page tracking process, we apply the coarse-to-fine matching technique as well as the adaptive keyframe-based tracking which updates the keyframe adequate for tracking as time goes by. The tracking process is motivated by Klein et al.'s work [12] which is well known as PTAM.

We explain details of the page recognition process in Section 3.1 and the page tracking process in Section 3.2.

3.1 Page Recognition Process

In the previous work, we assumed that a single page at most among the whole pages is visible. Therefore, all keypoints extracted from the entire image pass through the GRF and the page can be recognized using (1). However, the assumption is invalid in this work because our target pages in the image may be more than one. Our strategy to deal with this problem is to divide the image into subregions, inspect which pages are included in each subregion, and put together results using spatial locality. The left image of Fig. 3 shows the original image including two pages of which page IDs are 4 and 7 and the right image shows m x n subregions overlaid with keypoints.

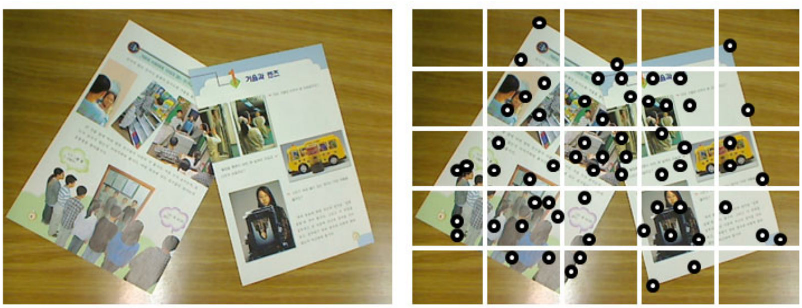


Fig. 3. (left) Original image, (right) Subregions overlaid with keypoints

We apply the GRF to each subregion individually in a way that only keypoints belonging to each subregion participate in calculating (1). Equation (1) is originally intended for averaging the probability distributions with respect to all keypoints and selecting the page with the highest probability, whereas we consider the top *T* pages instead of the only top one page. $Rank_r(i)$ is referred to as the ranking of the i-th page in the average probability distribution of the r-th subregion. To evaluate the likelihood that the r-th subregion might contain the i-th page, $Effect_r(i)$ is defined as (3).

$$Effect_r(i) = \begin{cases} T + 1 - Rank_r(i) & \text{if } Rank_r(i) \le T \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

If *Effect$_r$(i)* is greater than 0, we call the relationship between the r-th subregion and the i-th page "*the r-th subregion is effective to the i-th page*". Fig. 4 (a) shows the page IDs of which *Effect$_r$(i)* are greater than 0 in each region in decreasing order when T is 3. As shown in Fig. 4. (a), if the page ID actually comes from the true page, subregions with the same page ID have a tendency to gather together. We refer to this property as the spatial locality. Thus, if a large connected subregion group with the same ID is detected, the group is likely to include the page with the ID. *LCR(i)* is referred to as a set of 4-connected effective subregions related to i-th page. To evaluate the likelihood that the entire image might contain the i-th page, *Score(i)* is defined as (4).

$$Score(i) = \sum_{r \in LCR(i)} Effect_r(i) \tag{4}$$

Fig. 4. (b) and (c) show the top two connected subregion groups with the highest score, of which page IDs are 7 and 4. As shown in the both figures, it seems that the connected subregion group segments the real page region roughly.

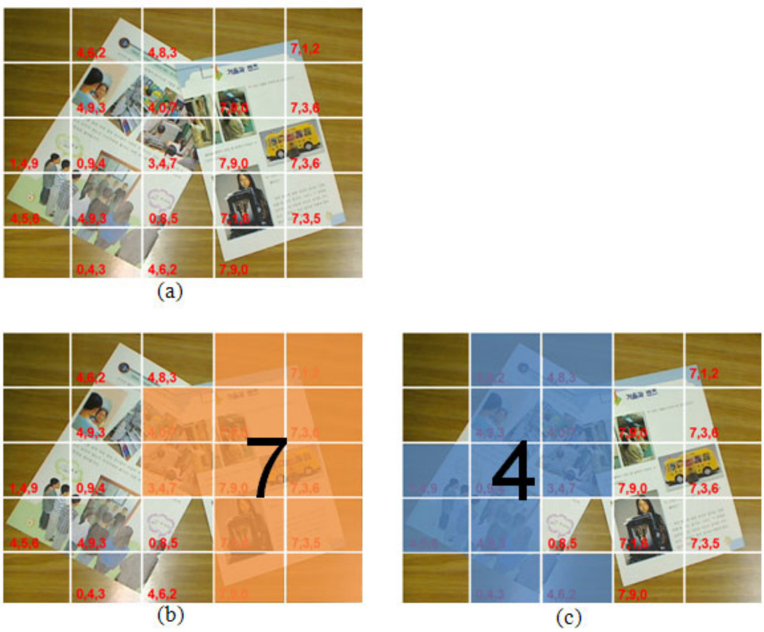


Fig. 4. Spatial locality of subregions

If pages which have already been recognized and tracked exist, it is not required to reconsider them in the page recognition process, so they are skipped. For all the newly recognized pages, keypoint matching is performed as in the previous work to calculate their initial poses. If the 7th page is recognized, only the keypoints belonging to

LCR(7) participate in the keypoint matching step using (2). There might exist outliers among the keypoint matching result, so it is required to remove the outliers. We employ the PROSAC method [10] for that purpose. Even if the page recognition results in wrong page IDs, the outlier removal step can filter them out. As the last step of the page recognition process, initial poses of recognized pages are calculated from the matching result by using [11]. The final page information P, a set of visible pages' information $p=(ID, R, t)$, is conveyed to the page tracking process and then poses are refined.

3.2 Page Tracking Process

When the new page is recognized by the page recognition process, we initially use the training image of the page as a keyframe to track the page. The pose of the page belonging to the training image has already been calculated in the training process. However, this is quite dangerous in terms of tracking stability because the training image could be taken from a different camera in different environments. Thus, we update the current frame as a keyframe just in case that it describes the page better than the existing keyframe according to the update score after every tracking success.

At every frame the following procedure is performed for tracking each page.

1) A prior pose is estimated from a motion model.
2) Map points in the world are projected into the image according to the estimated prior pose in step 1.
3) A coarse search is performed with 60 map points and the camera pose is refined.
4) A fine search is performed with at most 500 map points and the final pose is computed from the matching.
5) Update the motion model.

Camera motion M can be parameterized with a six-vector μ, which consists of three elements for translation and the remainder for rotation, using the exponential map [13]. Thus, given a camera pose P which transforms a point in a world coordinate into a point in a camera coordinate, the new camera pose $\hat{P}$ can be estimated as in (5) [12].

$$\hat{P} = M\,P = \exp(\mu)\,P\,, \tag{5}$$

where $P = [R\ t]$ and R and t are the camera rotation matrix and translation vector, respectively. The decaying velocity motion model is used which slows and stops eventually in case of the lack of new measurements.

To find matching pairs between map points in the world coordinate system and keypoints in a current image frame, a map point ($\boldsymbol{X}$) is projected into an image, as in (6).

$$x = K[R\ t]\boldsymbol{X} = K[R\ t]\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{6}$$

where x is a 2D point in an image coordinate and K is the intrinsic matrix of the camera.

We perform an affine warping to approximate viewpoint changes between the 8x8 image patch generated from the keyframe of the world map and the current camera position, as described in [12]. The determinant of the affine warping matrix is used to determine the pyramid level at which the patch could be searched. The best match between the projected map point and a keypoint in the current image frame can be found within a fixed radius around the projected map point position by evaluating zero-mean SSD scores within the circular search region.

To make the page tracking process more robust to rapid camera motions, patch search and pose update are done twice. First, a coarse search is done with only 60 map points from the highest levels of the image pyramid of the current frame. Patch search is performed with a larger radius and a pose is refined with the successful matching pairs, by minimizing the Tukey biweight objective function [14] of the reprojection error iteratively. With the refined pose, a fine search is done up to 500 map points. Now the patch search is performed with a smaller search region. The final camera pose is calculated eventually and the camera motion is updated from the difference between the initial and final camera pose of the frame.

A tracking would likely fail by a motion blur, occlusion, or an incorrect position estimate. Thus, if a fraction of the keypoint matching falls below a certain threshold, it is considered as a tracking failure and this page is eliminated from the tracked page list, so the page can be re-recognized through the page recognition process.

The tracking quality in the page tracking process depends on the quality of the keyframe because it is used in the patch search. However, because the initial keyframe of each page is from the offline stage and it could be captured from a different camera in different environments, the fraction of the keypoint matching is likely to fall, which might cause the unexpected tracking failure as well as a poor tracking quality. Thus, the goal of the keyframe update is to capture the image frame as a keyframe for the page which describes the page well, satisfying the following three conditions

1) An image is clear enough with no motion blur.
2) The area of a page appears as much as possible in the image and it is captured as large as possible in the image.
3) The page plane and camera center are orthogonal.

The total score function of the t-th frame is the weighted sum of the three sub score functions of $\mathrm{Score}_{\mathrm{ZMSSD}}$, $\mathrm{Score}_{\mathrm{area}}$, and $\mathrm{Score}_{\mathrm{ortho}}$, as shown in (7)

$$\begin{aligned} Score_{total}(I_t) = \; & \omega_1\, Score_{ZMSSD}(I_t) \\ & + \omega_2\, Score_{area}(I_t) \\ & + \omega_3\, Score_{ortho}(I_t)\,, \end{aligned} \tag{7}$$

where $\mathrm{Score}_{\mathrm{ZMSSD}}$, $\mathrm{Score}_{\mathrm{area}}$, and $\mathrm{Score}_{\mathrm{ortho}}$ represent the above conditions in sequence and ω_1, ω_2, and ω_3 are the weight factors of the three scores which represent their importance, respectively. $\mathrm{Score}_{\mathrm{ZMSSD}}$ measures how similar the adjacent frames are with no motion blur as in (8). ZMSSD is the zero-mean squared sum of distance between two adjacent blurred images at frame *t* (BI_t) and *t-1*(BI_{t-1}).

$$Score_{ZMSSD}(I_t) = 1 - \frac{ZMSSD(BI_t, BI_{t-1})}{ZMSSD\_max} \tag{8}$$

$\mathrm{Score}_{\mathrm{area}}$ measures how much portion of the pages are shown within the image at frame *t*, as in (9).

$$Score_{area}(I_t) = \frac{Area_t}{ImageSize} \frac{Area_t}{AreaOfPage_t}, \tag{9}$$

where $\mathrm{Area}_{\mathrm{t}}$ is the area of the page shown in the image at frame t in pixel scale and $\mathrm{AreaOfPage}_{\mathrm{t}}$ is the area of the page, including the area beyond the image boundary after projecting four boundary points of the page into the image according to the camera pose.

$\mathrm{Score}_{\mathrm{ortho}}$ measures how orthogonal the page is to the camera z vector, by calculating an inner product of the camera z vector ($\mathrm{CamZ}_{\mathrm{z,t}}$) and the inverse of the page normal vector ($-\mathrm{PageNorm}_{\mathrm{z,t}}$), as in (10).

$$Score_{ortho}(I_t) = CamZ_{z,t} \cdot (-PageNorm_{z,t}) \tag{10}$$

Therefore, the higher $\mathrm{Score}_{\mathrm{total}}$ is the more accurately the page tracking process tracks the pose of a page. Fig. 5. shows the progress of the keyframe update from (a) to (f), where (a) is the training image used as the initial keyframe; it is quite different from the other keyframes captured in real-time. Progressing from (b) to (f), it is ascertained that the keyframe is more orthogonal and fit to the image to be favorable for tracking.

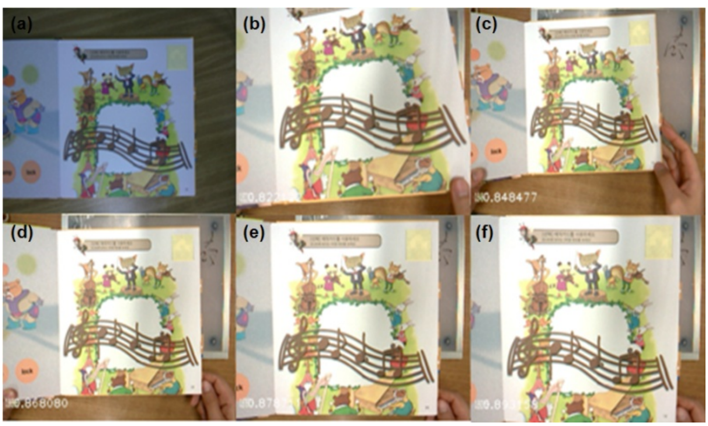


Fig. 5. The progress of the keyframe update from (a) to (f)

4 Experiment Results

For the experiment, a desktop of 3.17GHz Core 2 duo CPU with 2GB memory and a NVIDIA GeForce 8600 graphic card were used. A logitech Q9000 webcam was attached to the desktop. A 640x480 image was obtained from the camera. Although the dual core CPU was used in the experiment, the proposed method can be performed on a single core CPU because it uses a single thread.

The experiments were performed with a science textbook including 166 pages for elementary students. We took one picture of each page and extracted 250 keypoints per page. The keypoints were used in training the GRF with the number of trees $NT = 40$ and a depth of $d = 10$.

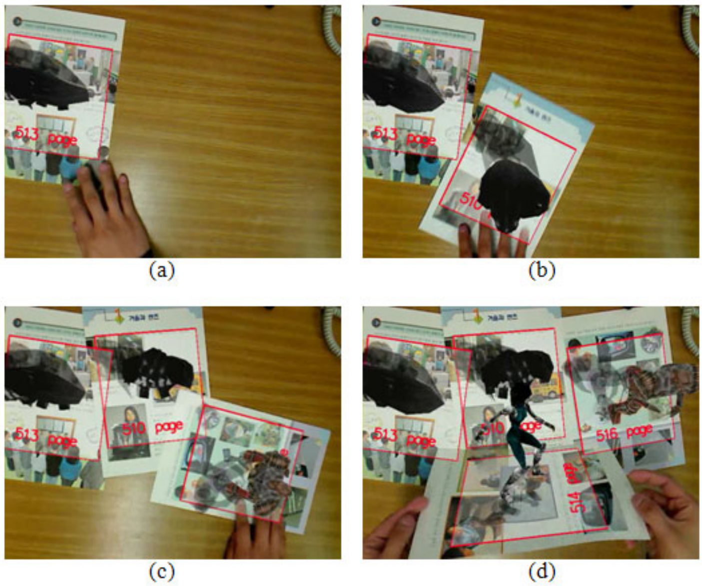


Fig. 6. Multiple Page Recognition and Tracking

In real-time, 300 keypoints are extracted from an image captured by the camera, the image is divided into 4 x 5 subregions and keypoints belonging to each subregion are passed through the GRF. We consider the top 10 page IDs in each subregion (T=10 in Equation (3)). Although the proposed method can recognize and track regardless of the number of visible pages, we impose a limitation to handle up to four pages due to the real-time constraint (more than 25 fps). Fig 6. (a)-(d) show that one to four pages are recognized and tracked.

Besides the ordinary conditions such as Fig. 6., multiple pages are tracked well in situations with dramatic viewpoint variation, scale variations, illumination variations, and partial occlusions thanks to the adaptive keyframe-based tracking, as shown in Fig. 7.

One of the most important things to decide a tracking quality is the jitter. To measure the jitter, we used the reprojection error as measurement and compared the adaptive keyframe based tracking with the fixed keyframe based tracking, where the training image was always used as the keyframe. Fig. 8. (a) shows a setup for measuring the jitter, where a test page and the camera did not move and 1000 frames were selected to measure the reprojection error.

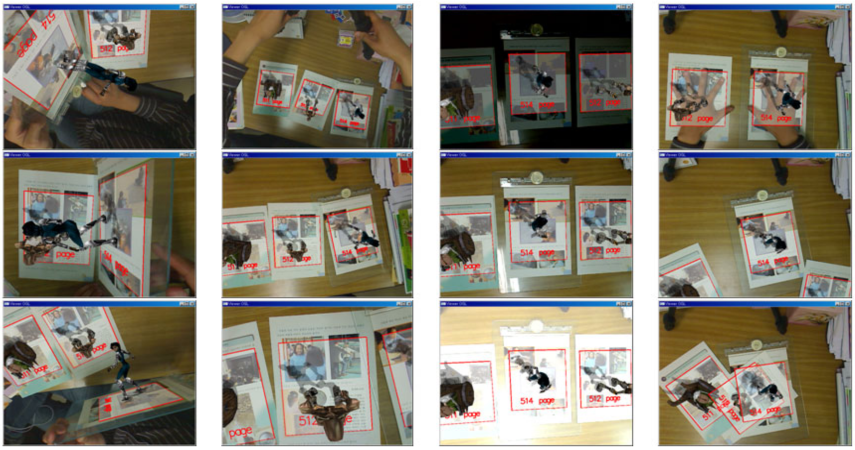


Fig. 7. Tracking in situations with dramatic viewpoint variation (1st column), scale variations (2nd column), illumination variations (3rd column), and partial occlusions (4th column)

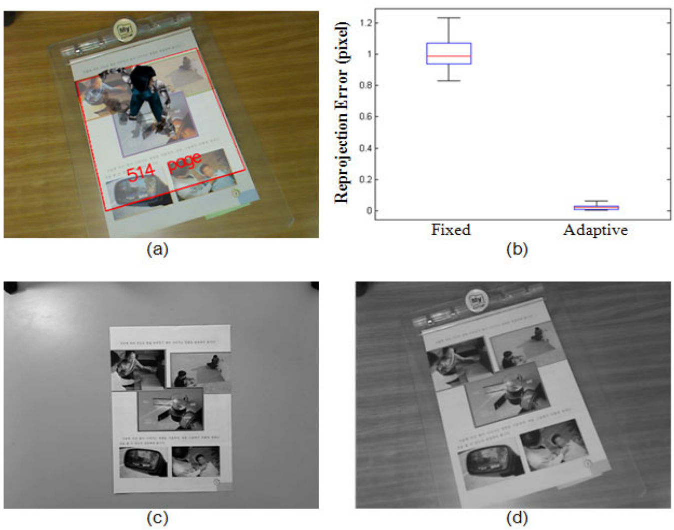


Fig. 8. (a) a setup for measuring the jitter, (b) the comparison result of the fixed and the adaptive methods, and the keyframes (c) for the fixed and (d) for the adaptive methods

Fig. 8. (b) shows the comparison result using a boxplot. The means of the fixed and the adaptive keyframe based tracking are 1.002 pixels and 0.026 pixels, respectively. This indicates that the adaptive method results in much less jitter than the fixed method. Fig. 8. (c) and (d) show the keyframes for the fixed and the adaptive methods.

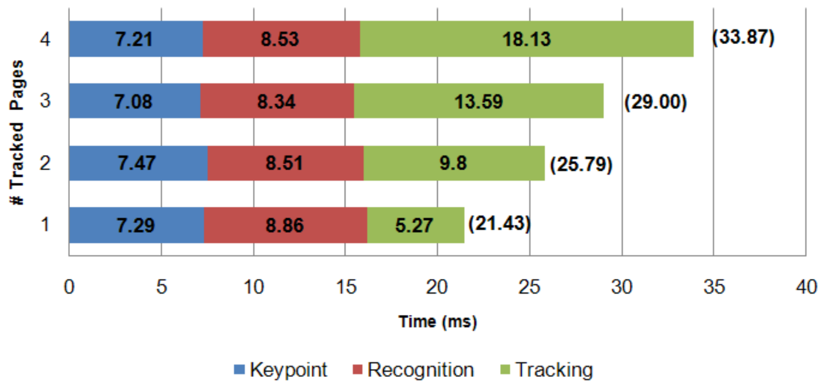


Fig. 9. The required times

Finally, we observed the required time as the number of tracked pages increases. Fig. 9. shows the average times for tracking in consecutive 100 frames according to the number of tracked pages. As expected, the more pages are tracked, the time is longer. The overall time consists of the keypoint extraction time, the recognition time, and the tracking time. The number of tracked pages has no effect on the times required for the keypoint extraction and recognition, but the tracking time increases linearly. If the number of pages tracked at a time is restricted to four, the proposed method can achieve approximately 30 fps.

5 Conclusion

For augmented books, this paper presents the markerless visual tracking method capable of recognizing and tracking multiple pages in real-time. The proposed method builds on our previous work using the generic randomized forest (GRF). The previous work finds out one page in the entire image using the GRF, whereas the proposed method detects multiple pages by dividing an image into subregions, applying the GRF to each subregion and discovering spatial locality from the GRF. We also propose the adaptive keyframe-based tracking which updates the keyframe adequate for tracking as time goes by.

As a result, the proposed method is robust to various situations with dramatic viewpoint variation, scale variations, illumination variations, and partial occlusions. Thanks to the adaptive keyframe-based tracking, the jitter can be much reduced. Although the required time increases linearly as the number of tracked pages increases, our method can achieve 30 fps if we impose a limitation to handle up to four pages.

In the future, we have a plan to perform more experiments related to the accuracy of page recognition and a page pose. In addition, our method requires the complementary relationship between the recognition and tracking processes. We expect that this makes our method more efficient and faster.

References

1. Billinghurst, M., Kato, H., Poupyrev, I.: The MagicBook: A Transitional AR Interface. Computers and Graphics, 745–753 (2001)
2. Cho, K.S., Lee, J.H., Lee, J.S., Yang, H.S.: A Realistic e-Learning System based on Mixed Reality. In: 13th International Conference on Virtual Systems and Multimedia (2007)
3. Fumihisa, S., Yusuke, Y., Koki, F., Toshio, S., Kenji, K., Asako, K., Hideyuki, T.: Vivid Encyclopedia: MR Pictorial Book of Insects. In: Proc. Virtual Reality Society of Japan Annual Conference (2004)
4. Taketa, N., Hayash, K., Kato, H., Nishida, S.: Virtual pop-up book based on augmented reality. In: Smith, M.J., Salvendy, G. (eds.) HCII 2007. LNCS, vol. 4558, pp. 475–484. Springer, Heidelberg (2007)
5. Scherrer, C., Pilet, V., Fua, P., Lepetit, V.: The haunted book. In: 7th IEEE/ACM International Symposium on Mixed and Augmented Reality (2008)
6. Kim, K., Park, J., Woo, W.: Marker-Less Tracking for Multi-layer Authoring in AR Books. In: 8th International Conference on Entertainment Computing, pp. 48–59 (2009)
7. Cho, K., Yoo, J., Yang, H.S.: Markerless Visual Tracking for Augmented Books. In: Joint Virtual Reality Conference EGVE-ICAT-EURO VR, pp. 13–20 (2009)
8. Lepetit, V., Fua, P.: Keypoint Recognition using Randomized trees. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(9), 1465–1479 (2006)
9. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
10. Chum, O., Matas, J.: Matching with PROSAC-Progressive Sample Consensus. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 220–226 (2005)
11. Klein, G.: Visual Tracking for Augmented Reality, PhD. Thesis, pp.165–169, University of Cambridge (2006)
12. Klein, G., Murray, D.: Parallel tracking and mapping for small AR workspaces. In: 6^{th}IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 255–234 (2007)
13. Varadarajan, V.: Lie Groups, Lie Algebras and Their Representation. Springer, Heidelberg (1974)
14. Huber, P.: Robust Statistics. Wiley, Chichester (1981)

Online Scene Modeling for Interactive AR Applications

Jaesang Yoo[1], Kyusung Cho[2], Jinki Jung[2], and Hyun S. Yang[2]

[1] Electronics and Telecommunications Research Institute, 138 Gajeongno, Yuseong-gu, Daejeon, 305-700, Republic of Korea
jsyoo@etri.re.kr

[2] Department of Computer Science, Korea Advance Institute of Science and Technology, 373-1 Guseong-dong, Yuseong-gu, Daejeon, 305-701, Republic of Korea
{qtboy,jk,hsyang}@paradise.kaist.ac.kr

Abstract. Augmented reality applications require 3D model of environment to provide even more realistic experience. Unfortunately, however, most of researches on 3D modeling have been restricted to an offline process up to now, which conflicts with characteristics of AR such as realtime and online experience. In addition, it is barely possible not only to generate 3D model of whole world in advance but also trasfer the burden of 3D model generation to a user, which limits the usability of AR. Thus, it is required to draw the 3D model generation to an online stage from an offline stage. In this paper, we propose an online scene modeling method to generate 3D model of a scene, based on the keyframe-based SLAM which supports AR experience even in an unknown scene by generating a map of 3D points. The scene modeling process in this paper is a little computationally expensive in terms of real-time but it doesn't restrict real-time property of AR because it is executed on a background process. Therefore, a user will be provided with an interactive AR applications that support interactions between the real and virtual world even in an unknown environment.

Keywords: Keyframe-based SLAM, higher level structure, online modeling, augmented reality.

1 Introduction

Recently Augmented Reality (AR) technology is gaining attention by the spread of mobile devices. The application of AR is expanding to Xbox Project Natal, PS3, and Nintendo DSi beyond the mobile platforms. By broadening the environment for AR applications, the importance of AR contents is getting bigger.

It is still difficult to create AR contents in spite of a growth on development tools and circumstances for AR applications. Simple applications such as Layar 2.0 and Sekai camera to develop AR contents have been released. However, it is hardly possible to build 3D models of arbitrary unknown scenes for AR applications and moreover it takes a lot of time to build a 3D model of a scene. Thus, we propose a method of an online scene modeling for interactive AR applications in real-time.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 139–150, 2010.

Recently many researches related to 3D modeling for interactive AR contents have been published. Bergig et al.'s work [1] has presented a 3D model authoring tool by hand sketching using normal pen and paper. Moreover Hengel et al.'s work [2] has presented a tool to easily generate 3D models from a video clip in which objects for modeling are appear. However, [1] has limitation because they are only capable of simple modeling. Also [2] is not fully online: the user should build models in offline.

In this paper, we propose an online scene modeling method to generate 3D model of an environment for interactive AR applications with minimum interaction with a user. The system requires a user to do nothing but moving the camera to capture the scene of which the user wants to generate a 3D model. This paper focused on an online scene modeling to recognize geometry of a scene and incorporate it into the system for interactive AR applications.

2 Related Work

2.1 Keyframe-Based SLAM

Simultaneous Localization and Mapping was begun in Robotics to localize a robot in an unknown environment. The early days, it used LRF and sonar sensors[4,5] but recently, vision sensor was also used [6,7]. The first stage of vision-based SLAM, only the sensor was substituted to a camera from LRF or sonar, keeping the overall framework of SLAM using probabilistic models such as EKF(Extended Kalman Filter) and particle filter. Davison et al.'s work [6] was based on EKF and constructed a map of 3D sparse points so that it contained uncertainties on the camera's position by nature. Later, PTAM(Parallel Tracking and Mapping) so-called keyframe-based SLAM appeared [7]. To localize a moving camera in an unknown scene, tracking and mapping are separated and run in two parallel threads. PTAM constructs a map of 3D points based on keyframes collected during tracking process using batch techniques, Bundle Adjustment and the map can contain thousands of points, which allows precise camera localization more proper for AR applications.

2.2 Scene Modeling

The detection of higher level structure in a real world has been progressed using both visual and non-visual sensors. Planes were fitted to clouds of points reconstructed using LRF(Laser Ranging Finder)[8,9]. The discovered planes were, however, not incorporated into a map.

Afterward, many have attempted to discover planes in vision-based SLAM as in [10,11,12]. Rachmielowski et al.'s work [10] attempted to get 3D reconstruction of a scene during SLAM tracking. They connected sparse 3D points to compose a mesh using a Delaunay triangulation and reconstructed a scene model with the meshes. Thus, the reconstruction is unlikely a good estimation of the real surface because the 3D points used for a mesh were sparse and involved uncertainties by nature of the SLAM approach. Chekhlov et al.'s work [11] also tried 3D reconstruction of a scene during SLAM tracking. However, to find planes they approached with RANSAC process instead of the Delaunay triangulation used in [10]. In view of using RANSAC process, our system is similar to their system but we go beyond their work by incorporating the discovered planes into the map and using them during the tracking process.

3 Online Plane Fitting for Interactive AR

A map of PTAM usually consists of thousands of 3D points and it is impossible to consider the all points to generate a plane hypothesis. Thus, we face a problem of a working set selection, which will be discussed in 3.1. A method how to discover a planar structure in a scene will be presented in 3.2 and how to incorporate the discovered plane into a map will be presented in 3.3.

3.1 Region of Interest

Because considering all 3D points results in an inaccurate result as well as takes a lot of time when a plane hypothesis is generated, the points for the plane detection need to be classified. We call the classified region as ROI(region of interest).

In Hengel et al.'s work [13], a user selected the ROI by dragging a mouse. This was possible in their system because it worked in offline, not online. In our system, a user can select a working set by moving a camera view to the direction in which the user wants to generate a model, which is even more proper process for an online modeling system.

3.2 Detection of Planar Structure

In this section, it is introduced how to discover planes existing in real world. At first, it is described how to discover a plane from a 3D point cloud within ROI(region of interest) using RANSAC process in 3.2.1. However, it doesn't guarantee the discovered plane using RANSAC is a real plane in the real world because RANSAC process only considers a geometric distribution of 3D points. We call the discovered plane at this stage as a geometric consistent plane. Therefore, it is required a process to verify whether or not the generated plane exists in reality. This verification process is introduced in section 3.2.2 by testing if the geometric consistent plane is also photometric consistent.

3.2.1 Plane Hypothesis Using RANSAC

To discover a potential planar structure using RANSAC process, a similar approach to that used in [12] is used in this paper.

Step 1 : RANSAC robust estimation

A minimum of three 3D points is required to generate a plane hypothesis. Assuming X_1, X_2, X_3 are three 3D points, an origin p_0 and a normal vector n of the plane hypothesis π including those three points can be generated as followings.

$$p_0 = X_1 \qquad n = c_1 \times c_2 \qquad (1)$$
$$\text{where} \quad c_1 = X_2 - X_1 \qquad c_2 = X_3 - X_1$$

As soon as a plane hypothesis is generated with the minimum three points, consensus for the hypothesis among other points is determined using a perpendicular distance from the plane. The perpendicular distance $\mathrm{d}_{\perp}$ for a 3D point X_i from the plane $\pi[p_0, n]$ can be calculated as a following.

$$d_{\perp} = (X_i - p_0) \cdot n \tag{2}$$

To minimize effects of the ouliers, we impose a criterion that the distance must be less than or equal to some threshold d_{max}, that is, although $d_{\perp} > d_{max}$, we set $d_{\perp} = d_{max}$.

Step 2 : Optimal estimation

To compute the best-fit plane from inliers S_i resulting from RANSAC process, we use a principal component analysis approach similar to [3].

The origin of the plane hypothesis is set to the mean of the inliers and normal vector is set to the eigenvector of $M^T M$ corresponding to the smallest eigenvalue where M is a $l \times 3$ matrix stacked with position vectors for l inlying points w.r.t the mean. The smallest eigenvalue gives a measure of quality of the fit because it is the variance of the inliers in the normal direction.

3.2.2 Photo-Consistent Plane

So far, given a map of point clouds, we introduced how to discover a plane using RANSAC. The mainly considered property is the geometric distribution of the 3D points so that it is often likely to result in false detections of planes as shown in (c) of Fig 1. Therefore, a verification process is required to remove the false detections and

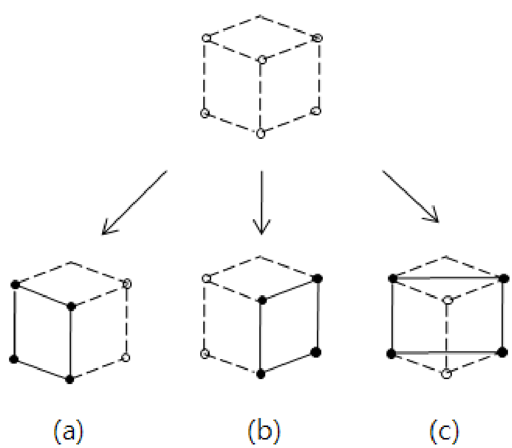


Fig. 1. Possibly detectable planes solely considered a geometric distribution of 3D points. (c) is false detection

the process called photo-consistency test is described below in details and summarized in algorithm 1.

Step 1 : Finding the visibility images $\mathcal{V}$

Given a set of keyframes I and a geometric consistent plane with geometric supports Π belonging to the plane, project 3D points within Π into every keyframe within I. If over 70 percent of overall points within Π appear on the keyframe, it is said that the plane is "visible" in the keyframe. To find a visibility image set $\mathcal{V}$, test if the plane is visible every keyframe.

$$\mathcal{V} \leftarrow \{i \mid \text{Ratio}(q, \Pi) > 0.7\} \quad (3)$$
$$\text{for } q \in \{\hat{q}_{i,j} \mid \hat{q}_{i,j} = P_i Q_j \text{ for } \forall\, i \in I, j \in \Pi\}$$

Step 2 : Computing the photo-consistency κ

Among visibility images $\mathcal{V}$, choose reference image I_{i_0} in which a plane appears dominant.

$$i_0 \leftarrow \text{argmax}_{i\in\mathcal{V}}\big(\text{Ratio}(q, \Pi)\big) \quad (4)$$
$$\text{for } q \in \{\hat{q}_{i,j} \mid \hat{q}_{i,j} = P_i Q_j \text{ for } \forall\, i \in I, j \in \Pi\}$$

Using 2D point correspondences between reference image I_{i_0} and the others in $\mathcal{V}$, compute homography $\text{H}_{i_0,i}$ satisfying the following relation.

$$\hat{q}_{i_0,j} \approx \text{H}_{i_0,i}\hat{q}_{i,j} \text{ for } \forall\, i \in \mathcal{V} \quad (5)$$

Once homography $\text{H}_{i_0,i}$ is computed, warp all images except reference one to the reference one.

$$\tilde{I}_i[q] \leftarrow I_i\big[\text{H}_{i_0,i}(q)\big] \quad (6)$$

$\tilde{I}_i$ is a warped image of image I_i to the reference one and q is a set of 2D points projected into an image I_i.

Ideally the warped image $\tilde{I}_i$ is identical to the reference image I_i However, because they are not identical due to image noises, we find the most suitable match in terms of a pixel value within radius r. Therefore, the photometric consistency score κ is induced by square root of the photometric error between the warped image and the reference one as a following.

$$\kappa^2(\pi) = \frac{1}{\#(\mathcal{V}-1)\#(q)} \sum_{i\in\mathcal{V}, i\neq i_0} \left(\sum_q \left(\min_{q', \|q'-q\|\leq r} \left|I_{i_0}[q] - \tilde{I}_i[q']\right|\right)^2\right) \quad (7)$$

If the photometric consistency score κ is greater than some threshold ε, reject the plane because it is not photometric consistent, that is, not exist in real world.

OBJECTIVE

Compute a photometric consistency κ for the plane with geometric support Π given the cloud of reconstructed 3D points Q_j, keyframe images I_i and cameras P_i and reject the plane if $\kappa(I_0) > \varepsilon$.

ALGORITHM

Step 1: Find the visibility images $\mathcal{V}$.

$$\mathcal{V} \leftarrow \{i \mid \text{Ratio}(q, \Pi) > 0.7\} \text{ for } q \in \{\hat{q}_{i,j} \mid \hat{q}_{i,j} = P_i Q_j \text{ for } \forall\, i \in I, j \in \Pi\}$$

where $\text{Ratio}(q, \Pi)$ is the ratio of the visibility measurements q to $\#(\Pi)$

Step 2: Compute the photometric consistency κ.

- Choose reference image i_0 :

$$i_0 \leftarrow \text{argmax}_{i \in \mathcal{V}}\big(\text{Ratio}(q, \Pi)\big)$$
$$\text{for } q \in \{\hat{q}_{i,j} \mid \hat{q}_{i,j} = P_i Q_j \text{ for } \forall\, i \in I, j \in \Pi\}$$

- Compute a homography $\text{H}_{i_0,i}$ induced by the point correspondences $\hat{q}_{i_0,j} \leftrightarrow \hat{q}_{i,j}$ such that

$$\hat{q}_{i_0,j} \approx \text{H}_{i_0,i}\hat{q}_{i,j} \text{ for } \forall\, i \in \mathcal{V}$$

- Warp all images I_i for $i \in \mathcal{V}$ and $i \neq i_0$ to the reference one :

$$\tilde{I}_i[q] \leftarrow I_i\big[\text{H}_{i_0,i}(q)\big] \text{ for } q \in \{\hat{q}_{i,j} \mid \hat{q}_{i,j} = P_i Q_j \text{ for } \forall\, i \in I, j \in \Pi\}$$

- Reject the plane if the consistency check fails, $i.e.$ if

$$\kappa(I_0) > \varepsilon,$$

where the consistency $\kappa(I_0)$ is given by

$$\kappa^2(I_0) = \frac{1}{\#(\mathcal{V}-1)\#(q)} \sum_{i \in \mathcal{V}, i \neq i_0} \left(\sum_{q \in I_0} \left(\min_{q', \|q'-q\| \leq r} \left| I_{i0}[q] - \tilde{I}_i[q'] \right| \right)^2 \right)$$

Algorithm 1. Photo-consistency test

3.3 Insertion of Planar Features

This section describes how to incorporate the discovered plane in 3.2 into the map of PTAM. The parameterization of a plane, addition of new point using a plane, and finally the way to reduce the computation using a plane are discussed in order.

3.3.1 Parameterization

To represent a plane, a center position and normal vector are necessary. They are 3 DOF each, and thus, a plane has total 6 DOF. This is over-parameterized into a transformation matrix $\boldsymbol{E}_{\mathcal{W}\mathcal{E}}$ from the local plane coordinate $\boldsymbol{\mathcal{E}}$ to the world coordinate $\boldsymbol{\mathcal{W}}$ with 12 internal parameters as in. A center of a plane is located on an origin and the normal vector is parallel to the z axis as in Fig 2. We set the x and y axis of the local plane coordinate parallel to the x and y axis of the world coordinate.

$$E_{\mathcal{W}\mathcal{E}} = \begin{bmatrix} R & t \\ \mathbf{0} & 1 \end{bmatrix} \tag{8}$$

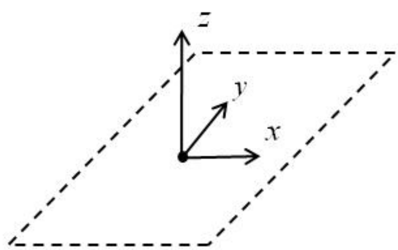


Fig. 2. Local plane coordinate $\boldsymbol{\mathcal{E}}$

3.3.2 New Mappoint Addition Using a Plane

To reconstruct 3D points, triangulation from correspondences between two views is used in PTAM. The number of iterations and final converging position of a bundle adjustment depend on the initial position of the reconstructed point. That is, the more accurate the initial position of the reconstructed point, the more accurate the adjusted position after the bundle adjustment as well as the faster the converging speed. Therefore we propose a method taking advantage of the plane information, homography, when a point is reconstructed as shown in Fig 3.

1. Find the closest keyframe to the new keyframe among keyframes in a map and set it as a target.
2. Find matching correspondences between new and target keyframes.
3. Compute a homography H_1 from the matching correspondences.
4. Compute H_2 from 2D point correspondences between the target keyframe and a plane coordinate.
5. With H_1 and H_2, the position on the plane coordinate p_{new} of a new point x_{new} is determined as below.

$$p_{new} = H_2 H_1 x_{new} \tag{3.12}$$

6. A 3D point on a plane X_{new} is reconstructed from a 2D point on a plane coordinate p_{new} using transformation matrix $E_{W\mathcal{E}}$ from the plane coordinate to the world coordinate.

$$X_{new} = E_{W\mathcal{E}}\, P_{new} \quad 3.9)$$

$$where\ \ p_{new} = (x, y)^{\mathrm{T}} \text{ and } P_{new} = (x, y, 0, 1)^{\mathrm{T}}$$

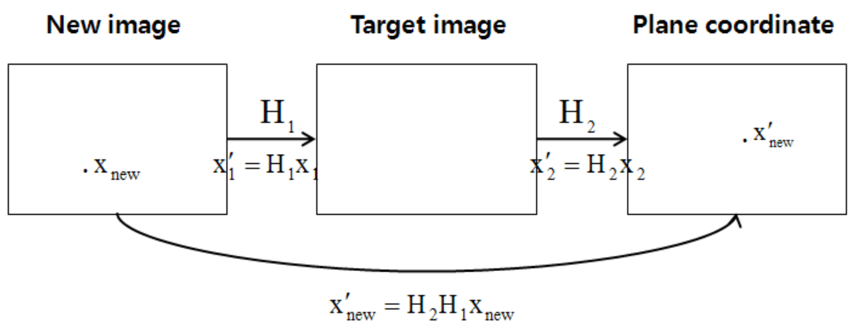


Fig. 3. A 3D point reconstruction using homography given a plane

3.3.3 Selective Measurement

Without any prior knowledge of a scene, all mappoints are used to estimate the pose of the camera. To be frank, thousands of 3D points are not necessary and a few hundred are enough to estimate the pose of the camera, using planes. As a result, it is likely to save the computational time by knowing the scene information and actually it is as showed in the experiment below.

4 Experiment

For the experiment, we use a laptop of 2.83 GHz Quad CPU with 2GB memory. Logitech Ultra webcam is attached to the laptop to obtain a 640 × 480 image.

Fig. 4. Man-made environment with three orthogonal planes

Experiments were performed in a man-made environment in which three orthogonal planes appeared with a lot of textures as in Fig 4.

Fig 5 shows how this system discovers and incorporates plane structures in a scene into PTAM map. Points in white indicate 3D reconstructed points from stereo analysis and colored points indicate points that belong to the discovered plane by method in this paper. Different color shows different planes. As a result three orthogonal planes are discovered and incorporated into the map.

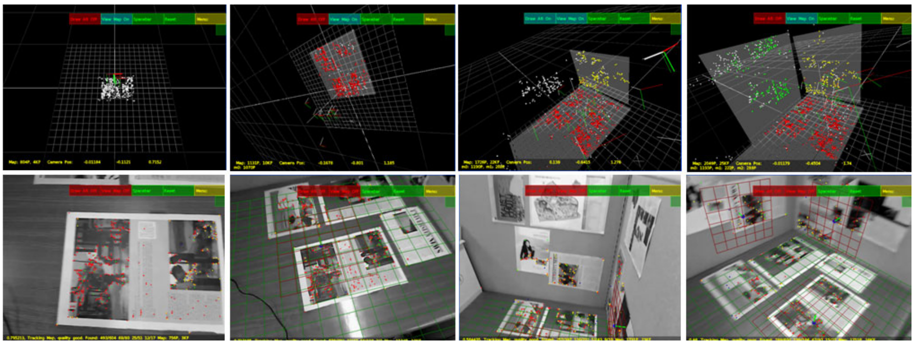


Fig. 5. Discovery of a scene geometry (left to right)

Fig 6 shows changes in a matching rate and time consumption as the number of measurements per a plane varies from 50 to 200 with video clip of 1800 frames. As a result, the test shows that tracking with higher level structures like a plane can potentially reduce the time consumption compared with the tracking only with points, resulting in similar tracking accuracy. As an example, with 150 points per a plane, tracking time consumed to track a camera in same video clip decreased 13ms to 9ms, which meant 30% reduction in time.

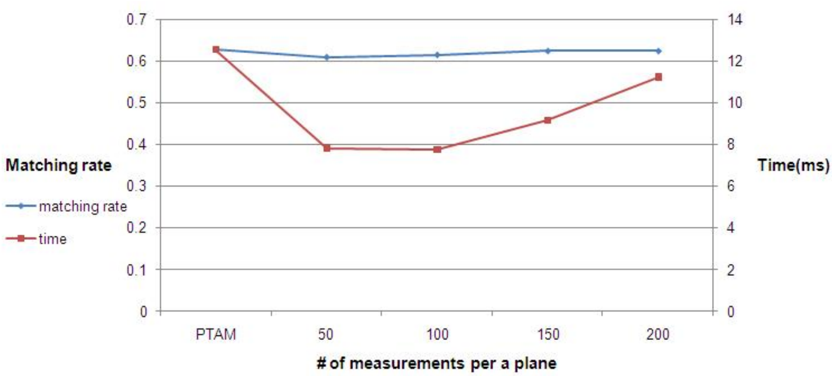


Fig. 6. Changes in a matching rate(blue, left axis) and tracking time(red, right axis) as the number of measurements per a plane varies

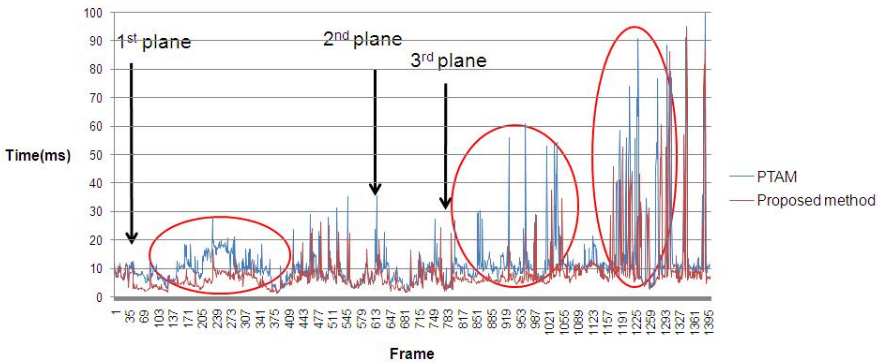


Fig. 7. Tracking time comparison of PTAM and proposed method of this paper on an identical video clip of 1400 frames

Fig. 8. A virtual gravity ball rebounding from real ground and walls

To compare proposed method with PTAM in terms of overall performance in time consumption, we plotted the tracking time of the PTAM(blue line) and proposed method(red line) on an identical video clip of 1400 frames in Fig 7. 150 points per a plane was chosen in the test and first, second, and third planes were discovered and incorporated into a map on 35th, 620th, and 780th frame. On the beginning of the test before 35 frames when first plane was discovered and incorporated into a map, both showed similar time consumption. However, since 35th frame, it showed the proposed method was superior to PTAM in terms of time consumption. The frames within red circles in Fig 7 indicated such superiority.

Fig 8 shows a simple interactive AR application that interacts with a real world. A white ball is a virtual gravity ball and it interacts with a scene, two walls and ground, by rebounding when it collides with them. A full video clip is attached with this paper.

5 Conclusion and Future Work

For interactive AR applications, this paper presented a method to discover and incorporate a higher level structure, a plane from point clouds on online. However, a plane is just a little bit higher level structure than a point cloud. There are a lot higher level structures than a plane in our environment where AR applications will set up. Therefore, for the future work, we need to discover and incorporate other higher level structures such as sphere, cube, cylinder, and polyhedral to describe a real world environment so that ultimately interactive AR application can be achieved. As a result of the future work, a real interaction not only with a plane such as a wall and ground but with objects such as a box, ball, cup and any other objects existing in real world can come true.

Acknowledgement

"This research is supported by the Ubiquitous Computing and Network (UCN) Project, Knowledge and Economy Frontier R&D Program of the Ministry of Knowledge Economy (MKE) in Korea and a result of subproject UCN 10C2-J2-11T".

References

1. Bergig, O., Hagbi, N., El-Sana, J., Billinghurst, M.: In-Place 3D Sketching for Authoring Mechanical Systems. In: 8th IEEE International Symposium on Mixed and Augmented Reality, pp. 87–94 (2009)
2. Hengel, A.V.D., Hill, R., Ward, W., Dick, A.: In Situ Image-based Modeling. In: 8th IEEE International Symposium on Mixed and Augmented Reality, pp. 107–110 (2009)
3. Weingarten, J., Gruener, G., Siegwart, R.: Probabilistic plane fitting in 3D and an application to robotic mapping. In: IEEE International Conference on Robotics and Automation, pp. 927–932 (2004)
4. Montemerlo, M., Thrun, S., Koller, D., Wegbreit, B.: FastSLAM: A Factored Solution to Simultaneous Localization and Mapping Problem. In: National Conference on Artificial Intelligence (2002)
5. Montemerlo, M., Thrun, S., Koller, D., Wegbreit, B.: FastSLAM 2.0: An Improved Particle Filtering Algorithm for Simultaneous Localization and Mapping that Provably Converges. In: International Joint Conference on Artificial Intelligence (2003)

6. Davison, A., Reid, I., Molton, N.D., Stasse, O.: MonoSLAM: Real-time single camera SLAM. IEEE Transaction on Pattern Analysis and Machine Intelligence 29(6), 1052–1067 (2007)
7. Klein, G., Murray, D.: Parallel Tracking and Mapping for Small AR Workspaces. In: 6th IEEE International Symposium on Mixed and Augmented Reality, pp. 225–234 (2007)
8. Nevado, M.M., Gercía-Bermejo, J.G., Zalama, E.: Obtaining 3D models of indoor environments with a mobile robot by estimating local surface directions. Robotics and Autonomous Systems 48(2-3), 131–143 (2004)
9. Viejo, D., Cazorla, M.: 3D plane-based egomotion for SLAM on semi-structured environment. In: IEEE International Conference on Intelligent Robots and Systems (2007)
10. Rachmielowski, A., Jägers, M., Cobzas, D.: Realtime visualization of monocular data for 3D reconstruction. In: Canadian Conference on Computer and Robot Vision, pp. 196–202 (2008)
11. Chekhlov, D., Gee, A., Calway, A., Mayol-Cuevas, W.: Ninja on a plane: Automatic Discovery of Physical Planes for Augmented Reality Using Visual Slam. In: 6th IEEE International Symposium on Mixed and Augmented Reality, pp. 1–4 (2007)
12. Gee, A., Chekhlov, D., Calway, A., Mayol-Cuevas, W.: Discovering Higher Level Structure in Visual SLAM. IEEE Transactions on Robotics, 980–990 (2008)
13. Hengel, A.V.D., Dick, A., Thormählen, T., Ward, B., Torr, P.H.S.: Building Models of Regular Scenes from Structure-and-Motion. In: Proc. 17th British Machine Vision Conference, pp. 197–206 (2006)

Unnecessary Image Pair Detection for a Large Scale Reconstruction

Jaekwang Lee[1] and Chang-Joon Park[2]

[1] Smart Game Research Team, Contents Research Division, Electronics and Telecommunications Research Institute (University of Science and Technology), Daejeon, Korea
[2] Smart Game Research Team, Contents Research Division, Electronics and Telecommunications Research Institute, Daejeon, Korea
ljk3815@naver.com, chjpark@etri.re.kr

Abstract. This paper proposes an algorithm to detect unnecessary image pairs for efficient structure from motion. Since image pair with small baseline is considered as a poor condition for reconstruction, we focus on computing cameras closely located. We address a term, "*remoteness*" which indicates the distance between two images in this paper. The *remoteness* is not affected by image's intrinsic parameters because camera intrinsic matrix is applied to put the extracted features in the normalized coordinate. The *remoteness* is computed using feature disparity in normalized coordinate. Therefore, we can detect redundant image pair captured at the near position without reconstruction. The proposed algorithm is proved by experimental results with Notre Dame images.

Keywords: web image collection, reconstruction, unnecessary image pair, remoteness.

1 Introduction

As digital cameras are generalized, we can easily get thousands of pictures of a certain tourist spot from internet. Many researchers have developed algorithms dealing with large number of images downloaded from internet. Snavely et al.[1], [2] propose a structure from motion algorithm using internet photo collections and Chum et al.[3], [4] propose a large scale web image clustering method by computing similarity. In order to compute similarity, a near duplicate image detection(NDID) approach[3]-[5] is used and it is applied to detect duplicated images for 3D reconstruction[1], [2] as well.

3D reconstruction means a construction of a three-dimensional model of an object from several two-dimensional views of it. In other words, the location of cameras is very important factor for reconstructing a scene. It is a common agreement that if two images are captured with wide baseline and many correspondences, then the images are good pair for reconstruction. However the near duplicate image detection method computes superficial similarity for every matching pair. Even though the algorithm works well for detecting image pairs which have similar image content, it is not a proper method to remove duplicate images for 3D reconstruction using NDID.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 151–159, 2010.

Snavely et al.[6] propose a novel method for a large scale reconstruction using skeletal graph. They reduce a computational complexity for the reconstruction by making uncertainty based skeletal graphs. They compute reconstruction and covariance for each matching pair and then remove near duplicate images. While removing the near duplicate images, they use a NDID method but it is not good enough to remove redundant images for 3D reconstruction because it detects similar images instead of the closely located images. However our method(we'd like to call it Nearly Located Image Detection method, NLID) not only is able to detect similar images but also is able to detect images with small baseline. It is helpful to reduce parameters and reconstruction cost. In this paper, an algorithm for detecting the nearest neighboring image by their locations are proposed. In order to extract the nearest camera, comparative distances are computed between one and the others. We define a term, "*remoteness*" as a measurement of how far away between two cameras are.

To make the algorithm be robust to these problems, the feature extraction method, SIFT[8] and the feature matching method, RANSAC[9] are employed in this paper. The main requirements for good results are that the extracted correspondences are accurate and the number of features should be enough to guarantee the two images containing common scene or object. We only consider the pair with more than 30 correspondences. Without loss of generality, we assume that every image has EXIF information as mentioned in [2]. From the EXIF tag, we can obtain the focal length, so the features for each image can be easily normalized by multiplying the inverse of calibration matrix to each feature[10]. We show that the proposed *remoteness* measures can be efficiently computed and used successfully to find the nearest camera in this paper.

The remainder of this paper is structured as follows. In section 2 we review the related works and a nearly located image detection method is described in section 3. In section 4 some primary results are shown to prove that the proposed algorithms can detect the nearest cameras successfully. We conclude with future work in section 5.

2 Previous Work

Camera localization is well studied in computer vision and robotics. Several algorithms have been proposed attempting to give camera locations by image search or structure from motion. Typically sparse features such as interest points and straight edges are detected and described. Restrictive assumptions such as static scenes[11] or the existence of 3D models[12] are used to match those visual features between query images and exemplar database images of the same scene viewed under different viewpoints or illumination conditions. In Simultaneous Localization and Mapping (SLAM), both camera motion and the 3D points are recovered[13].

Self-localization can be categorically divided into global and find localization[14]. The first one considered in this paper determines the position and orientation of the robot pose within a world coordinate system. The second one deals with the dynamic-pose of the robot. They conceptualized as the extraction and processing of image features, which by means of recursive state estimations provide the continuous location of the cameras. However, partially significant visual landmarks and assessed poses provide insufficient useful information for real applications. Also the height of the camera should be the same.

Johansson and Cipolla[12] have developed a system to automatically determine the 3D viewpoint position related to model buildings using an image. For each building, they create a template and map assuming linearity of the object in high contrast regions. Matching is then performed by measuring the fit of each 3D model to the query image. Sivic and Zisserman present video google[15] where key frames from movies are used to query and find similarities between them using a text retrieval approach. The descriptors extracted from the images are quantized into visual words. The collection of the visual words are then used in Term Frequency Inverse Document Frequency(TF-IDF) scoring. The scoring is accomplished using a list of references to the images. Although location of features may be of importance on small vocabularies, it becomes insignificant when the vocabulary grows. All these approaches tend to work well for recognizing specific locations but we should have reference image data.

3 Nearly Located Image Detection

In this section, a method finding *remoteness* between two cameras is described As mentioned above, we assume that the cameras are calibrated(i.e., focal length is given for each image) and features are matched completely.

The *remoteness* is obtained by the process as follows:

- Locate the feature points in normalized coordinate
- Center the converted feature points
- Select a point randomly
- Rotate all features to put the selected point on *x* axis
- Compute *remoteness* by differencing the points

3.1 Feature Points in Normalized Coordinate

The resolution of each image collection can be vary. To deal with features regardless of its resolution, the features should be in normalized coordinate. By applying inverse of the calibration matrix to the feature points, we can express the points with the points in normalized coordinate.

As described in [10], the calibration matrix K can be written as

$$K = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

where, a_x, a_y are the scale factor in the x and y coordinate direction, respectively, s is the skew, and $(x_0, y_0)^T$ are the coordinates of the principal point. Normally, we assume that the camera has zero skew, the pixels are square, and also the principal point is at the image centre. Then, the calibration matrix can be

$$K = \begin{bmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

where, f is the focal length of the image and x_0 and y_0 are the half of the image resolution.

By multiplying the inverse of calibration matrix, K to the features, the features are normalized and we can deal with the normalized features with regardless of the image's zooming effect(focal length) and size.

3.2 Center the Features in Normalized Coordinate

Let the normalized feature points for first and second camera are $\hat{p}$ and $\hat{q}$, respectively, then a centre point is computed by averaging the points.

$$\bar{p} = \frac{1}{n}\sum_{i=1}^{n}\hat{p}_i, \qquad \bar{q} = \frac{1}{n}\sum_{i=1}^{n}\hat{q}_i \tag{1}$$

and the centered points by

$$p' = \hat{p}_i - \bar{p}, \qquad q' = \hat{q}_i - \bar{q}$$

Note that the sum of the centroid points is zero.

3.3 Compute *remoteness*

In order to compute *remoteness* between two cameras, we should consider the fact that the images can be rotated at the random angle. By rotating the selected point to be on the x axis, two coordinate systems of features can be the same. By doing this, we can overcome the rotation problem of images.

We can rotate the feature points by multiplying the rotation matrix

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

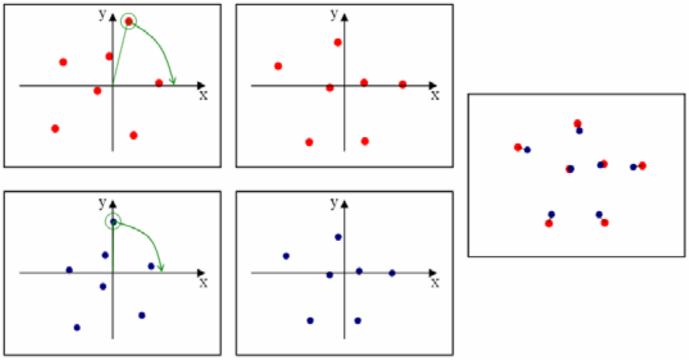
Where, $\cos\theta = \frac{x}{\sqrt{x^2+y^2}}$, $\sin\theta = \frac{-y}{\sqrt{x^2+y^2}}$ and x, y are the coordinates of the selected point. Then, rotated points can be denoted by

$$u_i = R \cdot p_i', \qquad v_i = R \cdot q_i' \tag{2}$$

Note that we ignore a scale factor to discriminate the camera position from the object. The *remoteness* can be represented by differencing the features

$$R = \frac{1}{n}\sum_{i=1}^{n} e_i = \frac{1}{n}\sum_{i=1}^{n}|u_i - v_i| \tag{3}$$

Fig. 1 depicts the process for computing the *remoteness*. The first row of Fig. 1 is for the first camera features, the second row is for the second camera and the axes in each image indicate image coordinate axes with origin (0, 0). Fig. 1(a) shows points



(a) Features in norm. coord. (b) Rotated features (c) Pixel difference

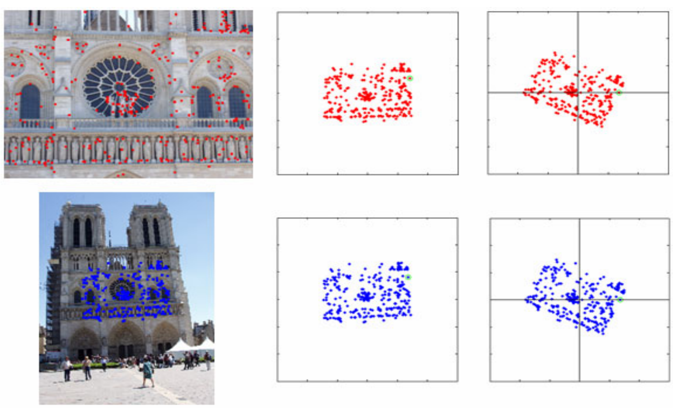
Fig. 1. Process to find the remoteness

normalized and centered, and randomly selected point and its correspondence to take rotation for each image feature. Fig. 1(b) is a result of rotated features to make the selected point lie on the x axis by eq(2). Fig. 1(c) is a disparity between normalized and rotated features of two images. To make the R be general, R should be divided by the number of correspondences.

4 Experimental Results

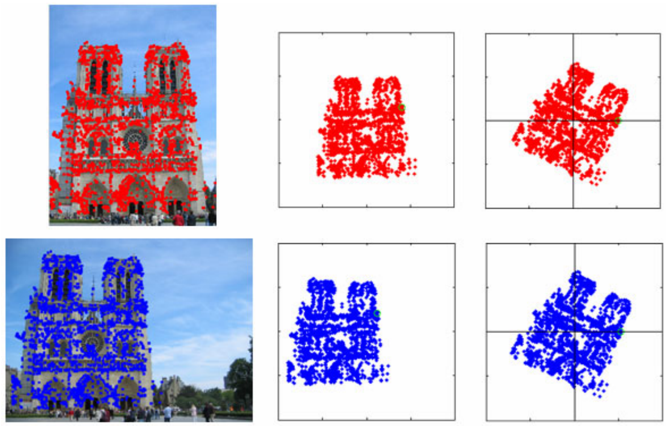
In this section, we apply our algorithm to various cases which can be happened in large scale images including zoom case.

We will not show the result for the cropped, resized and duplicated cases because we can easily see the *remoteness* is nearly zero for these cases.



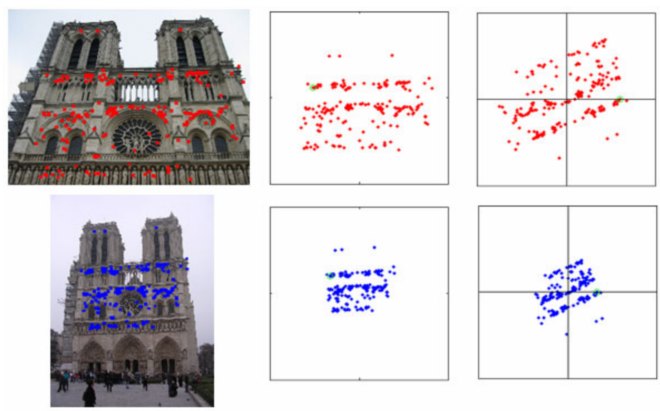
(a) original image+features (b) features in norm. coord. (c) rotated features

Fig. 2. *Remoteness* computation for a zoom in case



(a) original image+features (b) features in norm. coord. (c) rotated features

Fig. 3. *Remoteness* computation for a close case



(a) original image+features (b) features in norm. coord. (c) rotated features

Fig. 4. *Remoteness* computation for a distant case

Fig. 2 is a case to detect that the two images are taken at the same position. The *remoteness* of our method is 0.000304 with 316 features. Fig. 3 and Fig. 4 are the cases that the two images are taken at the near position and at the distant position. The *remoteness* of our method is 0.000561 with 2481 features and 0.1229 with 123 features. The small value of *remoteness* means they are captured at the near position and the pair is not good for reconstruction because of the small baseline.

Fig. 5 shows that the relationship between *remoteness* and distances with a selected image. The distance is computed after reconstructing the scene. The reason we reconstruct the scene is to demonstrate the relationship between them.

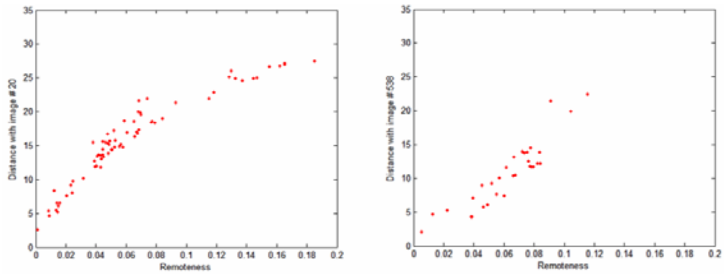


Fig. 5. *Remoteness* vs. distance with selected image

Fig. 6. Superimposed image with registration of cameras to an overhead map

Fig. 6 is a superimposed image of reconstructed camera locations to an overhead map. The reconstruction is performed to show whether our algorithm is good for detecting nearest neighboring camera or not. The figure depicts that the results of nearest cameras for 5 sample cameras. It shows that we can easily detect nearest camera without reconstruction, hence redundant image pairs for reconstruction are extracted simply.

5 Conclusions

In this paper we proposed algorithms to extract nearest cameras by their 3D location without reconstruction. By extracting nearest cameras we can easily remove unnecessary image pairs for 3D reconstruction. Since images contain EXIF information, the feature points can be located in normalized coordinate by applying inverse of calibration matrix. By differencing the points in normalized coordinate, *remoteness* is computed. Of course every pair has different number of correspondences, so we have to consider this factor for comparing *remoteness* with other image pair.

Recently most digital cameras have zoom lenses so practically many pictures downloaded from the internet are zoomed images even they are taken at the same location. In this paper, we showed that we can easily remove many redundant pairs including zooming, cropping and resizing cases using proposed algorithm. In addition we can partition the cameras by their location using Hierarchical clustering[7] which attempts to partition n images into nested clusters. In fact the *remoteness* is not a metric distance since it doesn't satisfy the triangle inequality. However the nearest camera pair can be uniquely determined so we can partition the cameras maintaining the closest pair.

By clustering the cameras we can make efficient graph like skeletal sets and eventually the reconstruction is simply performed. Hence efficient structure from motion can be achieved by using proposed algorithm and it can be applied for 3D scene reconstruction for various content applications such as virtual reality game, etc.

Acknowledgement

This work was supported by the Industrial Strategic Technology Development Program (KI002095, Online Game Quality Assurance Technology Development using Scenario Control of Massive Virtual User) funded by the Ministry of Knowledge Economy (MKE, Korea).

References

1. Snavely, N., Seits, S., Szeliski, R.: Photo Tourism: Exploring Photo Collections in 3D. In: ACM SIGGRAPH, pp. 835–846 (2006)
2. Snavely, N., Seits, S., Szeliski, R.: Modeling the World from Internet Photo Collections. Intl. Journal of Computer Vision 80(2), 189–210 (2008)
3. Chum O., Philbin J., Zisserman A.: Near Duplicate Image Detection: min-Hash and tf-idf Weighting. BMVC (2008)
4. Chum, O., Matas, J.: Web Scale Image Clustering. Technical Report CMP, CTU in Prague (2008)
5. Ke, Y., Sukthankar, R., Huston, L.: Efficient Near-duplicate Detection and Sub-image Retrieval. In: ACM Intl. Conf. on Multimedia (2004)
6. Snavely, N., Seits, S., Szeliski, R.: Skeletal Graphs for Efficient Structure from Motion. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1–8 (2008)
7. Eppstein, D.: Fast Hierarchical Clustering and Other Applications of Dynamic Closest Pairs. In: 9^{th} Symp. Discrete Algorithms, pp. 619–628 (1998)
8. Lowe, D.: Distinctive Image Features from Scale-invariant Keypoints. Intl. Journal of Computer Vision 60(2), 91–110 (2004)
9. Fischler, M.A., Bolles, R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartograph. Comm. of the ACM 24(6), 381–395 (1981)
10. Hartley, R., Zisserman, A.: Multiple View Geometry in computer vision. Cambridge university press, Cambridge (2003)
11. Wang, J., Zha, H., Cipolla, R.: Coarse-to-fine Vision-based Localization by Indexing Scale-invariant features. IEEE Trans. on Systems, Man and Cybernetics 36(2), 413–422 (2006)

12. Johansson, B., Cippola, R.: A System for Automatic Pose-estimation from a Single Image in a city scene. In: Intl. Conf. Signal Processing, Pattern Recognition and Applications (2002)
13. Williams, B., Klein, G., Reid, I.: Real-time Slam Relocalisation. In: Intl. Conf. on Computer Vision (2007)
14. Se, S., Lowe, D., Little, J.: Vision-based Global Localization and Mapping for Mobile Robots. IEEE Trans. on Robotics 21(3), 364–375 (2005)
15. Sivic, J., Zisserman, A.: Video Google: A Text Retrieval Approach to Object Matching in Videos. In: Intl. Conf. on Computer Vision, pp. 1470–1477 (2003)

Online Gaming Traffic Generator for Reproducing Gamer Behavior

Kwangsik Shin[1], Jinhyuk Kim[2], Kangmin Sohn[1], Changjoon Park[1], and Sangbang Choi[2]

[1] Smart Game Research Team, Content Research Division
Electronics and Telecommunications Research Institute,
138 Gajeongno, Yuseong-Daejeon, 305-700, Korea
{kwangsik,sogarian,chjpark}@etri.re.kr
[2] Electronic Engineering Department
Inha University
253 Yonghyun-dong, Nam-gu Incheon 402-751, Korea
graybaily@inha.edu, sangbang@inha.ac.kr

Abstract. In this paper, we proposed an online gaming traffic generator reflecting user behavior patterns. We analyzed the packet size and inter departure time distributions of a popular FPS game (Left4Dead) and MMORPG (World of Warcraft) for regenerating gaming traffic. The proposed traffic generator generates an inter departure time and gaming packet based on analytical model of the gamer behaviors, then transmits the packet according to the inter departure time. Packet generation results show that generated packets of World of Warcraft is much different with analytical model, unlike Left4Dead. It is caused by Nagle algorithm and Delayed Acknowledgments of TCP. Thus, we disabled the Nagle algorithm in the proposed traffic generator. The generation results show that the revised proposed traffic generator guarantees goodness of fit in the generated traffic distribution.

Keywords: Traffic generator, online gaming traffic, traffic analysis.

1 Introduction

With the recent global explosion of on-line multiplayer gaming, it is becoming more important to understand its network behavior and usage. Those are expected to be major contributors to World Wide Web traffic. On-line gaming is an increasingly popular form of entertainment on the Internet, with the on-line market predicted to be worth over $19.7 billion dollars in 2011 [1]. As an example of a popular, money-making game, World of Warcraft (WOW) has over 11,500,000 subscribers each paying a monthly fee in December 2008 [2]. Nowadays, in proposition to their success, the stability and performance of game servers become major issues in online games.

With the advances in computing and communication technologies, the issue of server testing using virtual user has been attracting interests from the researchers. These researches help reduce the costs and time required to test and deploy new applications and systems. The server testing technology using virtual clients has been

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 160–170, 2010.

studied very actively on file server, web server and etc [3-4]. The industry standard load testing solutions such as LoadRunner[5], QALoad[6] and e-Load[7], can emulate hundreds or thousands of concurrent users and prevent costly performance problems. Recently, online game server testing introduces the new research issues into the server testing technology [8-9]. These solutions enable to measure end-to-end performance, diagnose application, system bottlenecks and tune for better performance. One of the most important steps for these systems is the creation of tools for scalable, tunable, representative network traffic generation. Such tools are especially critical in emulation and simulation environments where representative background traffic is needed.

There are a number of works on the general topics of traffic characterization and modeling. Current studies for traffic generation tend to focus on either simple packet streams or recreation of a single application specific behavior [10-14]. However, interactive game traffic is much different with web or conventional internet traffic prevailing on the Internet today. It tends to employ small, highly periodic packets. Packets are small since the application requires extremely low latencies which make message aggregation and message retransmissions impractical. While conventional traffic generation tools are quite useful for evaluating the behavior and performance of host systems for traditional internet service, they are often inappropriate for use in online game tests. Therefore, in this paper, we propose an online gaming traffic generator reflecting user behavior patterns.

The rest of this paper is organized as follows. Section 2 summarizes related works for traffic generation. Section 3 provides a brief account of the online gaming traffic analysis results. Section 4 describes architecture of the proposed traffic generator and operation flow. Section 5 presents the result of traffic generation. And finally, section 6 draws some conclusions and future works.

2 Related Work

There are a number of works on the general topics of Internet traffic characterization and modeling. Current studies for traffic generation tend to focus on either simple packet streams or recreation of a single application specific behavior.

Packet streaming methods such as those used in tools like iperf [10] consist of sequences of packets separated by a constant interval. These methods form the basis for standard router performance tests such as those recommended in RFC 2544 and RFC 2889. Another example is an infinite FTP source, a commonly used traffic generator in simulations. While these approaches provide some insight into network system capabilities, they lack nearly all of the richness and diversity of packet streams observed in the live Internet [11]. Several successful application-specific workload generators have been developed as following. These generate application-level request sequences based on network traffic that has the same statistical properties as live traffic from the modeled application.

SURGE (Scalable URL Reference Generator) [12] is the tool for generating HTTP workload. One role for HTTP workload generation is as a means for understanding how servers and networks respond to variation in load. Empirical studies of operating Web servers have shown that they experience highly variable demands, which is exhibited as variability in CPU loads and number of open connections SURGE applies a number of observations like server file size distribution, request size distribution, relative file popularity, embedded file references, and idle periods of individual users of Web server usage to create a realistic Web workload generation tool which mimics a set of real users accessing a server.

Tmix [13] is a traffic generation system for the widely used ns-2 simulator. As the web have evolved, there remains no existing model of HTTP workloads that accounts for routine uses of the web for applications such as peer-to-peer file sharing and remote email access. In order to perform realistic network simulations, one needs a traffic generator that is capable of generating realistic synthetic traffic in a closed-loop fashion that "looks like" traffic found on an actual network. Tmix takes as input a packet header trace taken from a network link of interest. The trace is "reverse compiled" into a source-level characterization of each TCP connection present in the trace. The characterization, called a connection vector, is then used as input to an ns module called Tmix that emulates the socket-level behavior of the source application that created the corresponding connection in the trace.

D-ITG(Distributed Internet Traffic Generator) [14] is a platform capable to produce traffic that emulate sources of various protocols: TCP, UDP, ICMP, DNS, Telnet and VoIP (G.711, G.723, G.729, Voice Activity Detection, Compressed RTP). D-ITG produces traffic accurately adheres to patterns defined by the inter departure time between packets (IDT) and the packet size (PS) stochastic processes. Also, it can perform both one-way-delay (OWD) and round-trip-time (RTT) measurement, packet loss evaluation, jitter and throughput measurement. D-ITG improves log-server using distributed architecture especially for performance enhancement achieved by the sender (in terms of generated data rate) and the receiver (in terms of received data rate). Log-server is used by senders and receivers to maintain the information needed to compute statistics about the experiment made.

These typically focus on generating sequences of conventional Internet applications that result in network traffic that has the same statistical properties as live traffic from the modeled application. However, interactive game traffic will be much different with web or conventional internet traffic prevailing on the Internet today. It tends to employ small, highly periodic packets. Packets are highly periodic as a result of the game's dynamic requirement of frequent, predictable state updates amongst clients and servers. While these tools are quite useful for evaluating the behavior and performance of host systems for traditional internet service, they are often inappropriate for use in online game tests. Therefore, in this paper, we propose a method suitable for modeling and generating MMOG's traffic.

3 Traffic Analysis

In this section, we analyzed the packet size and inter departure time (IDT) for various games of different genres, since online gaming traffic has different characteristics according to the type of game. Among them, we chose two recent popular online games, Left4Dead (L4D) [15] and WOW. The L4D is a cooperative first-person shooter (FPS) game by Valve Corporation and the WOW is a massively multiplayer online role playing game (MMORPG) by Blizzard Entertainment. FPS games provide large-scale gaming, and sometimes team-based combat in a real-time virtual environment. This is made through games utilizing the character's point of view, transmitting data from the client to the server, and then processing immediately the data received at that time. MMORPGs are a genre of computer role-playing games in which a very large number of players interact with one another within a virtual world. They focus on an accurate execution of client inputs, which has an impact on the transport protocol used. In the WOW, TCP serves as the transport protocol. TCP is connection oriented and offers reliability. This attribute is well suited for MMORPGs, preventing error propagation during long sessions.

Fig. 1 and 2 show distributions of IDT and packet size for the L4D and WOW, respectively. From the figures, we know that there are clear distinctions between two genres. In the fig. 1, mostly IDTs of the L4D are less than 50ms, while there are lots of packets with long IDT for the WOW. As a FPS game, the L4D imposes the hardest real-time requirements on a network. Since it is very sensitive to interactivity, this type of game requires low-latency point-to-point communication as well as directed broadcast channels to facilitate its real-time game logic. Therefore, packets are sent via UDP since clients should send packets at an interval that is much shorter than the time it would take to retransmit lost packets. In the other hand, fig. 2 shows that packet sizes of the WOW are less than ones of the L4D. Even though the MMORPG uses a TCP connection with long latency caused by error recovery mechanism, it reduces the round trip time (RTT) by using small packets.

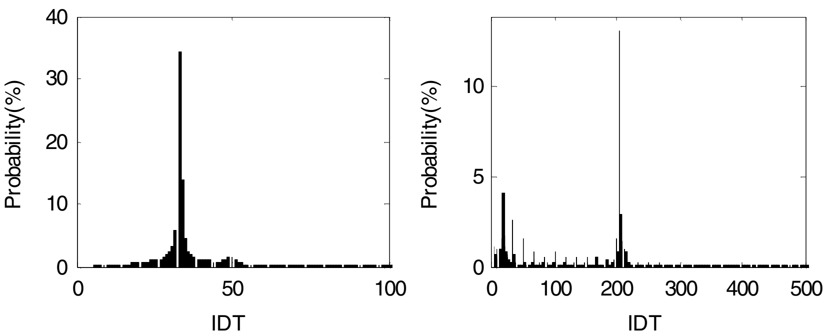


Fig. 1. IDT histograms for @ left) Left4Dead and @ right) WOW gaming packets

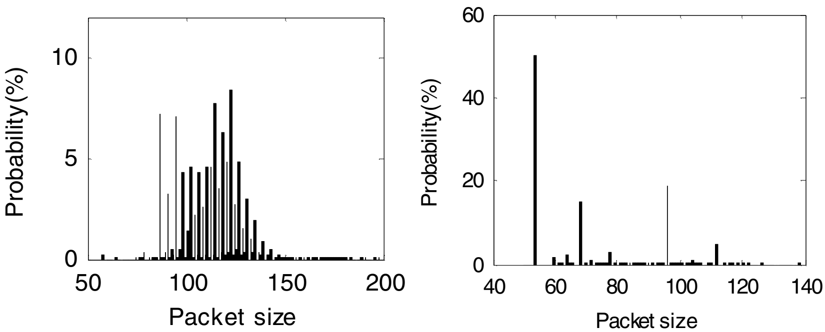


Fig. 2. Packet size histograms for @ left) Left4Dead and @ right) WOW gaming packets

While there are some differences according to game genre, online gaming traffic generally tends to employ small, highly-periodic packets in common. An in-depth knowledge of their traffic behavior will certainly assist network game developers and publishers to provide services and design their game servers better. In this paper, we propose an online gaming traffic generator in order to make available traffic characterizations of popular network games.

4 Architecture

In this section, we introduce the architecture of the proposed online gaming traffic generator shown in fig. 3. In the figure, *Analytical Model* is a result of analysis for each online game. Since online gaming traffic is typically messy data that is very difficult to analyze, we already studied an analytical model designing method for gaming traffic in a preliminary work [16].

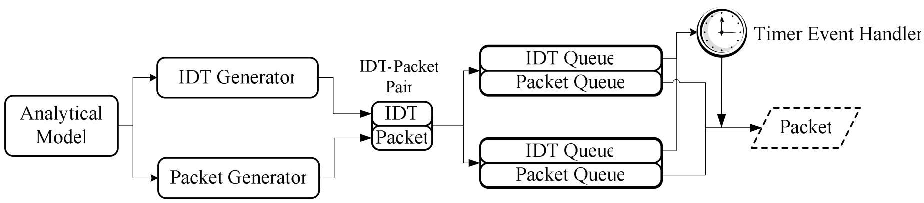


Fig. 3. The architecture of the proposed traffic generator

Packet Generator spawns a specified gaming packet classified by packet size and stores it on *Packet Queue*. The *Packet Queue* is implemented by double buffer to hold a block of data, so that *Timer Event Handler* will see a complete version of the data, rather than a partially-updated version of the data being created by *Packet Generator*. *IDT Generator* produces an interval, *IDT*, for triggering a *Timer Event Handler* that pulls out a generated packet. A *Timer Event Handler* is activated when an *IDT* is expired, and it takes out an *IDT* from the *IDT Queue* whenever it terminates execution. Since the *IDT Generator* is a producer of the *IDT Queue* and the *Timer Event Handler* is a consumer of it, we also implemented the *IDT Queue* as a double buffer. Because transmission of each generated packet is controlled by corresponding *IDT*, *Packet Queue* is synchronized with *IDT Queue* by *IDT-Packet Pair*.

Fig. 4 shows a flow chart of the proposed traffic generator. There are two threads; *Generator* and *Transmitter*. The former is a producer thread and the other is a consumer thread of *IDT & Packet Queue*. The producer initially finds a *Write_Q* and then verifies whether it is empty or not. If not, this process is repeated until detecting empty queue. In the other case, it generates an *IDT* and *Packet* based on *Analytical Model* of the gamer behaviors, and then groups them together prior to enqueue. After buffering *IDT-Packet Pair*, generating, pairing, and enqueueing processes are reiterated until corresponding queue is full. The consumer firstly gets a *Read_Q* and then checks state of queue. If the queue is not empty, it dequeues an *IDT-Packet Pair* and activates *Timer Event Handler*. The *Handler* sets timer as the *IDT* of dequeued

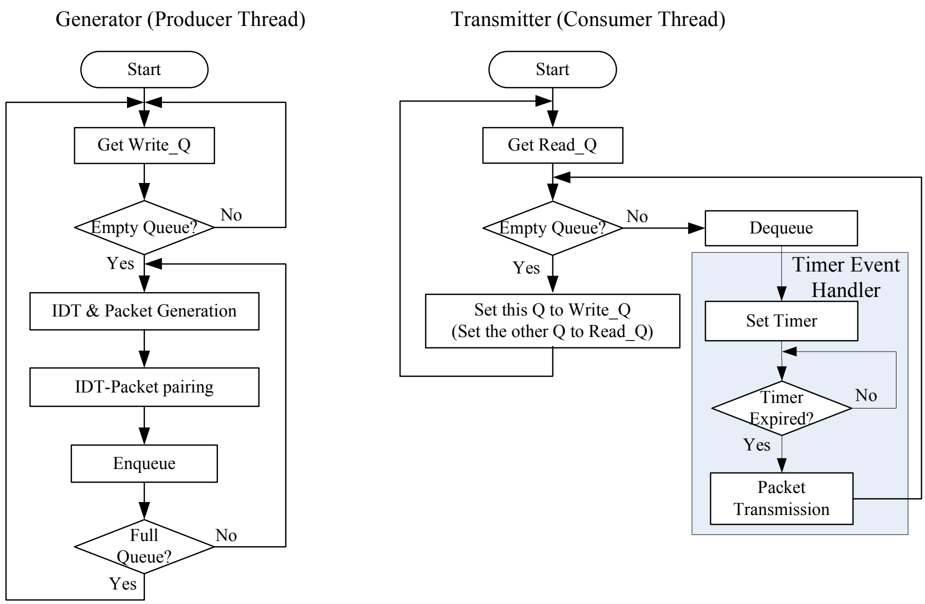


Fig. 4. Flow charts for (a left) Generator and (a right) Transmitter threads of the proposed scheme

IDT-Packet Pair and transmits the *Packet* of it after timer expiration. These procedures are continuously performed again until the queue is empty. When detecting unfilled queue, the *transmitter* defines corresponding queue as *Write_Q* and the other queue as *Read_Q*, respectively.

5 Evaluation

In this section, we compared generation results with analytical model based on previous analyzed gaming traffic. We captured produced packets of L4D and WOW at the receiver side and analyzed them into distributions of IDT and packet size as performed in the previous section. We show an intuitive comparison between the expected and generated histograms in fig. 5 and 6. In the figures, all expected histograms were drawn by analytical models for L4D or WOW. Fig. 5 shows that distribution graphs for both IDT and packet sizes of L4D are similar to ones of analytical model.

In spite of traffic generated by analytical model for IDT and packet size of WOW in fig. 6, however, there is a large discrepancy between generated packets and analytical model. It is caused by Nagle algorithm and Delayed Acknowledgments of TCP. Nagle algorithm is a means of improving the efficiency of TCP networks by reducing the number of packets that need to be sent over the network. It works by combining a number of small outgoing messages, and sending them all at once. In the other hand, Delayed Acknowledgement is used to reduce the number of packets sent on transmission media. If the socket layer immediately acked every packet, this would result in a lot of wasted acks, as usually the application sends a response shortly after receiving

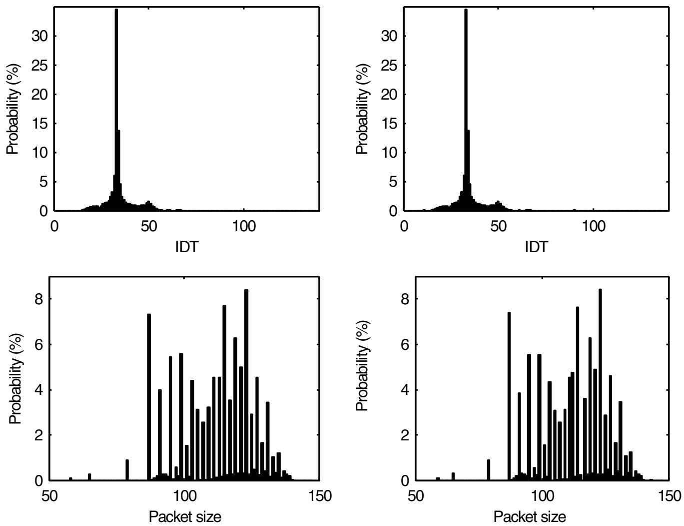


Fig. 5. Histograms for L4D; @ left-top) expected distribution for IDT, @ right-top) distribution of generated packets for IDT, @ left-bottom) expected distribution for packet size, and @ right-bottom) distribution of generated packets for packet size.

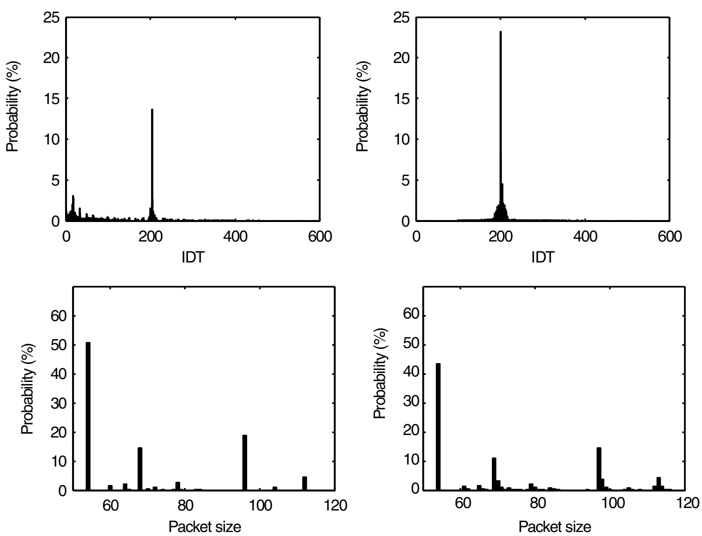
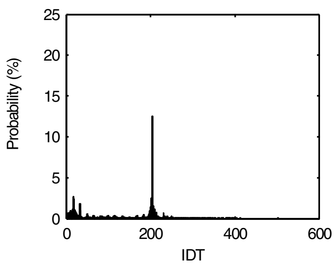


Fig. 6. Histograms for WOW; (@ left-top) expected distribution for IDT, @ right-top) distribution of generated packets for IDT, @ left-bottom) expected distribution for packet size, @ right-bottom) distribution of generated packets for packet size.

something, so TCP would be issuing two acknowledgments, one from the socket layer, one from the application layer. So the socket layer waits one or two hundred milliseconds, and if there is still nothing to send, sends an ack. Due to these techniques, even though traffic generator schedules packet transmission according to analytical model, TCP stack prevents some packets from leaving for a short time. The fig. 6 shows IDTs of generated packets are delayed until about 200ms and some small packets are united into one. Comparing with analytical model, ratio of small packets is reduced and some bigger sized packets are emerged. In this example, protocol header of WOW consists of TCP, IP, and Ethernet headers and its length is 54bytes.

From the fig. 6, we know applications expecting real time responses can react poorly with Nagle algorithm. Thus, we disabled the Nagle algorithm by setting the `TCP_NODELAY` socket option. Fig. 7 shows histogram for distributions of packets generated without Nagle algorithm. After removal of Nagle algorithm, distributions of generated packet are close to ones of analytical model.



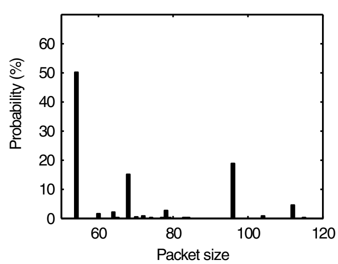


Fig. 7. Histograms for distribution of packets generated without Nagle algorithm; ⓐ left) IDT of WOW and ⓐ right) packet size of WOW

Finally, we displayed Q-Q plots of each model for a more detailed comparison. It is commonly useful in order to examine whether a generated dataset fits well an expected distribution. On this plot, the corresponding quartiles of both expected and generated distributions are graphed against each other, so that the deviations may be easily identified. If the two distributions being compared are similar, the points in the Q-Q plot will approximately lie on the line y = x. Fig. 8 shows Q-Q plots for the IDT and packet size of generated traffic.

It shows in detail which parts fit together appropriately. The Q-Q plot shows a pronounced difference among distributions of packets generated for L4D and WOW before and after disabling Nagle algorithm. From the fig. 8 (top) and (bottom), we know that the revised proposed traffic generator guarantees goodness of fit in the measured traffic distribution. Even though we inactivated Nagle algorithm, IDT of generated WOW is relatively worse than ones of generated L4D. Since the TCP provides various traffic control mechanisms, for example, a congestion control, flow control service, Delay Acknowledgement, and etc. as well as Nagle algorithm, arranged IDT of the traffic generator is distorted by TCP.

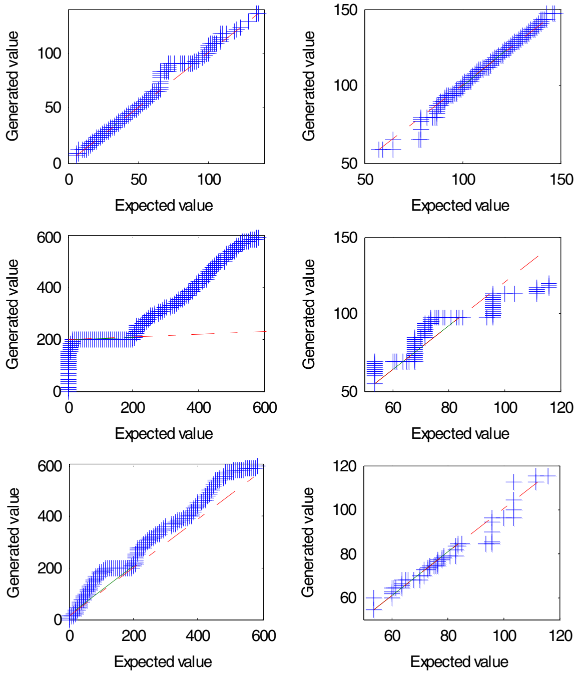


Fig. 8. Q-Q plots for distributions of generated packets; ⓐ left - top) IDT of L4D, ⓐ right - top) packet size of L4D, ⓐ left – mid) IDT of WOW, ⓐ right – mid) packet size of WOW, ⓐ left – bottom) IDT of WOW without Nagle algorithm, and ⓐ right – bottom) packet size of WOW without Nagle algorithm.

6 Conclusion and Future Work

In this paper, we proposed an online gaming traffic generator reflecting user behavior patterns. We analyzed the packet size and IDT for various games of different genres and introduced traffic characterizations of popular network games. The proposed traffic generator generates an IDT and packet based on analytical model of the gamer behaviors, and then transmits the packet according to the IDT.

In order to evaluate the proposed traffic generator, we compared generation results with analytical model based on previous analyzed gaming traffic. The results showed that generated packets of WOW is much different with analytical model while distribution graphs for both IDT and packet sizes of L4D are similar to ones of analytical model. It is caused by Nagle algorithm and Delayed Acknowledgments of TCP. Thus, we disabled the Nagle algorithm by setting the `TCP_NODELAY` socket option. The results show that the revised traffic generator guarantees goodness of fit in the generated traffic distribution. However, arranged IDT of the traffic generator is distorted by TCP stack, since the TCP provides various traffic control mechanisms, for example, a congestion control, flow control service, Delay Acknowledgement, and etc. as well as Nagle algorithm.

The problem of distortion by TCP stack remains an open research issue in gaming traffic generator. In order to improve accuracy for TCP based online game, we need to design delay model for various transmission control mechanisms of TCP. As a different approach, for bypassing TCP stack, we plan to implement traffic generator with a raw socket allowing direct sending and receiving of network packets by applications. In addition, we plan to measure workload of online game server according to massive gaming traffic generation.

Acknowledgement

This work was supported by the Industrial Strategic Technology Development Program (KI002095, Online Game Quality Assurance Technology Development using Scenario Control of Massive Virtual User) funded by the Ministry of Knowledge Economy (MKE, Korea).

References

[1] Korea Creative Contents Agency: 2009 White Paper on Korean Games, pp. 538–543 (September 2009)
[2] Blizzard Entertainment, `http://us.blizzard.com/`
[3] Arneson, D., Beth, S., Tavakley, R., Ruwart, T.: A Test Bed for a High-Performance File Server. In: Proceedings, Twelfth IEEE Symposium on Mass Storage Systems, pp. 26–29 (April 1993)
[4] Elbaum, S., Karre, S., Rothermel, G.: Improving web application testing with user session data. In: Proceedings. 25th International Conference on Software Engineering, pp. 49–59 (May 2003)
[5] LoadRunner, `http://www.hp.com`
[6] QALoad, `http://www.compuware.com/`
[7] Empirix, OneSight, `http://www.empirix.com/`
[8] Jung, Y., Lim, B., Sim, K., Lee, H., Park, I., Chung, J., Lee, J.: VENUS: The Online Game Simulator Using Massively Virtual Clients. In: Baik, D.-K. (ed.) AsiaSim 2004. LNCS (LNAI), vol. 3398, pp. 589–596. Springer, Heidelberg (2005)
[9] Cho, C., Sohn, K., Park, C., Kang, J.: Online Game Testing Using Scenario-based Control of Massive Virtual Users. In: 12th International Conference on Advanced Communication Technology, vol. 2, pp. 1676–1680 (February 2010)
[10] The iperf TCP/UDP Bandwidth Measurement Tool, `http://dast.nlanr.net/Projects/Iperf`
[11] Sommers, J., Kim, H., Barford, P.: Harpoon: a flow-level traffic generator for router and network tests. In: Proceedings of the Joint International Conference on Measurement and Modeling of Computer Systems, pp. 392–392 (June 2004)
[12] Barford, P., Crovella, M.: Generating representative Web workloads for network and server performance evaluation. In: Proceedings of the 1998 ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems, pp. 151–160 (June 1998)

[13] Weigle, M., Adurthi, P., Hernández-Campos, F., Jeffay, K., Smith, F.: Tmix: a tool for generating realistic TCP application workloads in ns-2. SIGCOMM Comput. Commun. Rev. 36, 65–76 (2006)
[14] Analysis and Experimentation of an Open Distributed Platform for Synthetic Traffic Generation. In: Proceedings of the 10th IEEE International Workshop on Future Trends of Distributed Computing Systems, pp. 277-283 (May 2004)
[15] Left4Dead, http://www.l4d.com
[16] Shin, K., Kim, J., Sohn, K., Park, C., Choi, S.: Transformation Approach to Model Online Gaming Traffic (unpublished)
[17] MSDN, http://msdn.microsoft.com

Click or Strike: Realistic versus Standard Game Controls in Violent Video Games and Their Effects on Aggression

André Melzer[1], Ingmar Derks[2], Jens Heydekorn[3], and Georges Steffgen[1]

[1] Université du Luxembourg, Campus Walferdange, L-7201 Walferdange, Luxembourg,
{andre.melzer,georges.steffgen}@uni.lu
[2] Bigpoint GmbH, D-20251 Hamburg, Germany
i.derks@bigpoint.net
[3] Otto-von-Guericke-Universität, Institut für Simulation und Graphik,
D-39106 Magdeburg, Germany
jheyde@isg.cs.uni-magdeburg.de

Abstract. The motion detection technology used in innovative game controlling devices like the Nintendo *Wii-Remote*® provides experiences of realistic and immersive game play. In the present study (N=62) it was tested whether this technology may also provoke stronger aggression-related effects than standard forms of interaction (i.e., keyboard and mouse). With the aid of a gesture recognition algorithm, a violent action role-playing game was developed to compare different modes of interaction within an otherwise identical game environment. In the *Embodied Gestures* condition participants performed realistic striking movements that caused the virtual character to attack and kill other in-game characters with a club or sword. In the *Standard Interaction* condition attacks resulted from simple mouse clicks. After the game session, participants showed a similar increase in negative feelings in both groups. When provided with ambiguous scenarios, however, participants in the *Embodied Gestures* condition tended to show more hostile cognitions (i.e., anger) than the *Standard Interaction* group. Results further corroborate the complexity of aggression-related effects in violent video games, especially with respect to situational factors like realistic game controls.

Keywords: Wii-Remote®, motion detection technology, gesture recognition, realistic interaction, violent video games, aggression, experiment.

1 Introduction

Video games have come a long way since their first steps some decades ago. Alongside general technological advances in computer science like, for example, improvements in graphics, so have video games developed into technologically advanced entertainment products. One particular aspect in the evolutionary process of human-computer interaction is the corresponding input and output hardware used in video games. For example, current game consoles like Nintendo *Wii*® combine accelerometer-based motion detection technology and force feedback functions, which aim at

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 171–182, 2010.

providing more direct or immersive gaming experiences. The upcoming Sony *Move* follows a related technological approach, whereas Microsoft's Xbox 360 add-on *Kinect* ("Project NATAL") does not require a controller but tracks whole body gestures and records spoken commands via a sensor device.

In the present study it was empirically tested whether technology realizing embodied gestures (e.g., *Wii-Remote*® control) may also affect psychological parameters more strongly than is the case with standard forms of interaction (i.e., keyboard and mouse)[1]. In particular, we were interested in aggression-related effects of violent video games: does performing realistic striking movements aimed at killing in-game opponents lead to greater levels of aggression than simply clicking a mouse button for the same purpose? To address this, a game was designed that allows for comparing different forms of interaction within an otherwise identical game environment. The next section describes the theoretical background of this approach.

2 Design and Implementation of a Test Game Environment

The design of a test game environment revolved around the idea of combining *Wii-Remote*® controls and keyboard/mouse controls in a single game. Commercial *Wii* games typically make exclusive usage of the *Wii-Remote*® and cannot be altered. However, there are several existing methods and toolkits for integration of that controller in other platforms than the *Wii*. Here we used a Bluetooth adapter to connect the *Wii-Remote*® to a *Windows*® PC and the C#/.NET/XNA programming environment for our implementations.

2.1 Game Analysis and Gesture Design

For this study we modified the existing open source game *Dungeon Quest*[2]. This game is an action role-playing game based on physical aggression directed at other in-game characters. The player navigates a virtual character through an underground maze and is forced to kill other humanoid in-game characters by beating them with a club or sword to find a hidden treasure.

The game got several game mechanical modifications. In particular we enhanced it by gesture recognition and support for *Wii-Remote*® usage.In the original version of *Dungeon Quest* the player controls the virtual character with keyboard and mouse. Right and left mouse clicks cause two different hitting attacks against other in-game characters. These character attacks were mapped onto two corresponding hitting gestures performed by the player (Fig. 1). Each gesture consists of two rapid successive movements. Performing them feels natural and is relatively simple, which is in line with common gesture recognition criteria [1]. Therefore, we developed a game concept, in which the primary interaction (hitting opponents) can either be initiated by conventional mouse clicks or by mimicking the character's attack animation, moving the *Wii-Remote*® device (which required implementation).

---

[1] The *Wii-Remote*® controlling device was used because information on this interaction device is widely available compared to similar devices respectively approaches.

[2] XnaProjects, Dungeon Quest GDC, http://xnaprojects.exdream.com/Default.aspx?Name=Dungeon%20Quest%20GDC, June 2010.

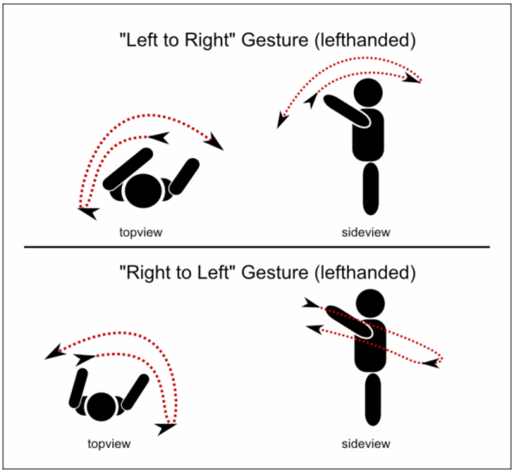


Fig. 1. Two hitting attacks in *Dungeon Quest* were mapped onto corresponding hitting gestures for the *Wii-Remote*® control.

2.2 Implementation

The game was set up in a way that allows controlling of game functions either by a combination of mouse and keyboard or by using the *Wii-Remote*®. This includes actual attacks as well as moving around, switching weapons or allocating skill points. We used the additional buttons of the *Wii-Remote*® to accomplish independency of the standard input devices when playing the game with the Nintendo device.

Based on a thorough analysis of gesture recognition algorithms in computer science, the Dynamic Time Warping (DTW) approach was selected for implementation. This algorithm has been used in various fields, including speech and visual pattern recognition (cf. [2], [19]). The program code of *Dungeon Quest* had to be modified to integrate the DTW algorithm for recording, filtering and recognition of movement data from the controller. Templates for two reference gestures were recorded (section 2.1) by performing the hitting gestures. Using templates on real performed gesture ensures the natural quality of reference patterns. Therefore, the algorithm uses these gesture templates to decide whether or not a hitting movement performed with the controller is an eligible attack gesture.

3 Psychological Effects of Video Games

Psychological studies on media effects have tested various aspects of video games. Positive effects have been demonstrated for educational games and entertainment games. With regard to spatial skills, for example, video games have been shown to facilitate spatial perception and spatial visualization (cf. [3]).

Generally, however, literature on video game effects more often than not refers to negative consequences, which have mostly been analyzed with regard to the impact of violent game content. Having said this, numerous studies tested whether or not game violence leads to an increase in real aggression[3]. There is ample evidence in the literature that exposure and active use of violent video games has negative effects on several dimensions. Compared to nonviolent games violent video games increase negative affect, aggressive thoughts, physiological arousal, and aggressive behavior, but reduce pro-social behavior in the players (e.g., [5]).

The General Aggression Model (GAM, [6]) may serve as a theoretical framework for the effects of violent video games. According to GAM, aggressive behavior results from either short-term or cumulated long-term internal processes. These may be triggered by person factors and situational factors. However, simply being exposed to situational factors like violent video games does not directly cause aggression. Rather, interacting internal state variables like arousal, feelings and thoughts mediate the effects. Compared to nonviolent games, for example, violent video games are known to make aggressive concepts more accessible, resulting in faster identification of aggressively connoted words [7] or greater availability of aggressive thoughts [8].

4 Experimental Study

In contrast to research that typically focuses on game content, the present study tested the effects of violent video game exposure from a *technologically* oriented perspective: does the technology-based simulation of violent behavior increase aggressive thoughts, feelings, and even cultivate enactive bodily representations that may be recoded into aggressive gaming behavior? We were especially interested in the moderating role of the type of I/O-device used when playing violent video games. I/O-devices directly affect the interactivity of video games; a hallmark of this medium, which also makes it unique among other media [9]. Interactivity provides the player with a bi-directional, contingent flow of information between the actively involved player and the feedback of the game's underlying software. Psychologically, interactivity facilitates experiences of causality and self-efficacy [10].

To date, only few studies have addressed the role of technological form dimensions in game effects, and knowledge is still sparse. Compared to older games, for example, technological advancement (i.e., newer games) was shown to increase players' sense of presence, involvement, and arousal [11]. In addition, virtual reality (VR) provides higher levels of involvement and immersion compared to video games that were played using a standard desktop interface [12]. These findings are of particular importance for research on media violence, because involvement and immersion are known to favor the player's *identification* with a violently acting character—an important factor that is associated with learned aggression [13].

Are there any differences in psychological effects of video games with regard to I/O-devices? Does playing a violent game with a motion detection device cause more aggression than utilizing keyboard and mouse? To our knowledge, only one study directly tested this but reported zero findings [14]; playing the violent video game *Manhunt 2* utilizing the *Wii-Remote*® did not lead to more aggressive cognitions than using a traditional game controller for the same game.

---

[3] Typically, aggression is defined as any behavior that is intended to harm another individual, whereas the target individual is motivated to avoid the harm [4].

Of course, it is premature to dispel concerns about the potentially negative effects of motion detection controls based on a single study. For example, Markey and Scherer [14] tested cognitive effects, but not behavioral measures players' affect. Compared to nonviolent games, violent VR games were found to increase aggressive feelings, game success, and fun [12]. Apparently, thus, technological form dimensions affect both psychological factors (i.e., emotion and cognition) and objective parameters (i.e., game success). The present study therefore tested whether motion detection controls make violent in-game behavior more efficient and lead to greater levels of negative affect and aggressive cognition than standard forms of interaction.

4.1 Method

To address these hypotheses, a violent action role-playing game called *Wii Remote Dungeon Quest* was developed that allows for comparing the effects of different modes of interaction within an otherwise identical game environment. In a third person view, the player navigates a virtual character through an underground maze to find a hidden treasure. Participants either performed realistic striking movements with the *Wii-Remote*® causing the virtual character to attack other in-game characters with a club or sword, or, when playing in the *Standard Interaction* condition, they initiated attacks by pointing and clicking (section 2.1). In addition, an automatic logging function in *Wii Remote Dungeon Quest* rendered objective parameters *how* the game is played. These parameters include *"number of opponents killed"* and *"total score"*. The game strategy participants pursued was also automatically recorded: *skill points* during the game may be invested into *Attack*, *Defense*, or *Speed* that significantly changes the behavior of the character controlled by the player.

4.1.1 Design and Participants

A one-factorial design was used. *Type of interaction* served as a between-subjects variable. Thus, participants played the violent video game *Wii Remote Dungeon Quest* either performing gestures using a *Wii-Remote*® (henceforth called *Gesture* condition) or keyboard and mouse (*Standard Interaction* condition).

A total of 62 students from different faculties at the University of Luxembourg took part in the experiment. In each of the two conditions, 16 females and 15 males participated, who were naïve to the experiment. The overall mean age was 21.33 (standard deviation [*SD*] = 2.12). Students were paid 5 Euros for their participation.

4.1.2 Materials

In this section, the game used in the present study will be described (4.1.2.1), together with the measuring instruments for the dependent psychological variables (4.1.2.2).

4.1.2.1 Wii Remote Dungeon Quest

Wii Remote Dungeon Quest provides two different modes of game control; *Wii Remote*® or keyboard and mouse. With the help of the additional *configurator* tool several parameters may be adjusted. Depending on the setting, for example, gesture recognition may either be run with the unmodified reference gesture or the mirrored gesture, thus allowing both left-handed and right-handed users to play the game.

An important parameter of *Wii Remote Dungeon Quest* is the tolerance (i.e., the threshold values) of the gesture recognition. Tolerance may be set to values as low as 500 and as high as 3,000, with the latter requiring just a slight shake of the wrist to be sufficient to initiate an attack. It was determined through testing that, on average, well-performed gestures resulted in a comparison value of about 1,500 or less (cf. section 2.2). This value required participants to perform arm movements using the *Wii-Remote*®, which are similar to realistic striking movements. Hence, a value of 1,500 was set as the default value for both gestures in the present study.

4.1.2.2 Measuring Instruments

The Positive and Negative Affect Schedule (PANAS, [15]) was used to measure positive and negative emotional experience. Each of the two dimensions comprises ten emotional adjectives (e.g., active, proud, afraid, hostile). We used the PANAS to assess relatively short-term fluctuations in mood (i.e., state affect). Therefore, participants rated each item on a 5-point scale ranging from 1=*very slightly or not at all* to 5=*extremely* reflecting the extent to which they experienced the emotion at the very moment (1) they entered the experimental lab, and (2) directly after the game. Internal consistency measures (α_1=.80 and α_2=.85) were within the expected range.

The so-called vignette technique was used to measure cognitive effects of violent video games. Participants were presented with four ambiguous conflict scenarios. Situations might be interpreted either as accidental or as intended acts of aggression. For example, participants were asked to imagine themselves being bumped into by another person carrying a cup of coffee resulting in large coffee stains on their white shirt. Participants rated each scenario on dimensions of anger (*"I feel angry"*), hostile attribution (*"the person did it on purpose"*), and desire for revenge (*"want to pay her back"*) on a 5-point scale ranging from 1=*no* to 5=*yes* (α=.74-.80).

It would be grossly negligent to rely on short-term indicators of aggression without paying attention to long-term characteristics that might moderate the results. Hence, individual trait aggression was measured using scales for anger (six items, e.g., *"At times I feel like a bomb ready to explode",* α=.77) and physical aggression (eight items, e.g., *"If somebody hits me, I hit back",* α=.75) from the Anger and Aggression Questionnaire [16]. Higher ratings on the 5-point scale indicated stronger affirmation.

4.1.3 Procedure

The experiment comprised, in chronological order, participants' ratings of their emotional state at the very moment they entered the lab, the training session in which participants learned how to utilize game controls, the study session in which they played *Wii Remote Dungeon Quest*, participants' ratings of their emotional state directly after playing, and the concluding questionnaire session that comprised demographical factors and ambiguous conflict scenarios. To control for potential sequential effects, half of the participants filled in trait aggression before game session whereas the other half received the trait measure after the game session.

Up to two participants were tested at the same time. By drawing a slip of paper they assigned themselves either to the *Gesture* condition or the *Standard Interaction* condition. They were then told that they would participate in testing a novel game. Neither the violent nature of the game, nor our interest in its effects was mentioned.

Fig. 2. Playing *Wii Remote Dungeon Quest* in the *Standard Interaction* condition using keyboard and mouse (*left*). Using the *Wii-Remote*® control in the *Gesture* condition (*right*) to perform striking movements causes corresponding hits performed by the in-game character.

Next, participants indicated their current positive and negative emotions by filling in the PANAS. They were then guided to a Dell PC equipped with Sennheiser HD212 pro headphones and a 22-inch Dell LCD display. In the *Gesture* condition, a slideshow explained how to use the *Wii-Remote*® control. In particular, the slideshow introduced two striking movements that cause the in-game character to perform the corresponding attack (section 2.1). In the *Standard Interaction* condition, instructions were presented how to play the game with keyboard and mouse. Attacks were initiated by pressing either the left or right mouse button. Participants were allowed five minutes to make themselves familiar with controls before they played the game for 15 minutes (Fig. 2). Immediately after the game session, participants filled in the PANAS again and provided details on demographical factors (e.g., age, gender, level of perceived game violence). They were then presented with the conflict scenarios and the trait aggression scales. The entire experiment took 40 to 50 minutes.

4.2 Results

Participants' trait aggression in the two experimental conditions will be described first (4.2.1). Next, analyses on the effects of type of interaction will be reported; participants' game performance (4.2.2), emotional experiences (4.2.3), and state aggression (i.e., cognitive effect; 4.2.4) will be compared between the *Gesture* and the *Standard Interaction* condition. Finally, additional findings will be presented (4.2.5).

4.2.1 Trait Aggression

The two experimental conditions were compared based on mean scorings in the trait aggression scales. As expected, two-sided *t*-tests (α=.05) revealed that there was no group difference with regard to trait anger ($M_{Gesture}$=2.49, SD=0.74; $M_{Standard\ Interaction}$=2.61, SD=0.82; p=.55) or trait physical aggression ($M_{Gesture}$=1.90, SD=0.49; $M_{Standard\ Interaction}$=1.80, SD=0.77; p=.54). Therefore, any differences in state aggression between groups are not reflecting stable individual characteristics. Rather, differences have to be attributed to influences caused by the type of interaction.

4.2.2 Game Performance: Objective Parameters

Analyses of game success showed that, in general, participants in the *Standard Interaction* condition were unexpectedly more successful than in the *Gesture* condition, $t(60)=2.41$, $p=.02$ (Table 1). This was also reflected by the number of opponents killed by their in-game character, $t(60)=3.59$, $p<.01$.

Participants in the *Standard Interaction* condition showed a greater tendency to strengthen their character's defense skills than their colleagues in the *Gesture* condition, $t(60)=1.95$, $p=.06$. With regard to attack and speed, however, both conditions showed similar patterns of game strategy, $ps \geq .39$.

Table 1. Mean overall score, number of opponents killed and number of skill points allocated to the in-game characters' attack, speed, or defense skills in the two experimental conditions.

	Gesture condition		***Standard Interaction***	
	M	*SD*	***M***	*SD*
Overall score	**962.71**	355.67	**1216.06**	17.22
# of opponents killed	**54.48**	464.70	**75.77**	28.23
Attack	**3.71**	3.48	**5.45**	4.84
Speed	**1.16**	2.34	**1.03**	1.64
Defense	**3.23**	4.14	**5.45**	4.84

4.2.3 Emotional Experiences: PANAS

Apparently, playing *Wii Remote Dungeon Quest* did not affect participants' level of *positive* emotions ($M=3.00$, $SD=0.77$) compared to their reports when they had entered the lab ($M=3.10$, $SD=0.60$), $t(59)=1.00$, $p=.32$. In contrast, *negative* emotions were clearly affected; after playing the violent game ($M=1.61$, $SD=0.58$) participants felt significantly worse than before ($M=1.31$, $SD=0.41$), $t(59)=-4.38$, $p<.01$. Remarkably, this was true for both experimental groups; participants in the *Standard Interaction* condition ($M=1.72$, $SD=0.63$) and in the *Gesture* condition ($M=1.48$, $SD=0.48$) reported almost similar levels of negative feelings, $t(60)=1.70$, $p=.09$.

4.2.4 State Aggression: Cognitive Effects

Participants' responses to conflict scenarios were separately calculated with regard to anger, hostile attribution, and desire for revenge (Table 2). Participants in the *Gesture* condition tended to respond to ambiguous scenarios with greater anger than participants in the *Standard Interaction* group, $t(60)=-1.49$, $p=.06$. Although group means were in the same direction both *hostile attribution* ($t=1.50$) and *desire for revenge* ($t=1.08$) failed to reach the level of significance, $ps \geq .14$.

Table 2. Mean responses to ambiguous conflict scenarios with regard to anger, hostile attribution, and desire for revenge in the two experimental conditions.

	Gesture condition		***Standard Interaction***	
	M	*SD*	***M***	*SD*
Anger	**3.54**	0.75	**3.14**	0.88
Hostile Attribution	**2.85**	0.71	**2.56**	0.81
Desire for Revenge	**2.10**	0.92	**1.84**	0.96

4.2.5 Additional Findings

Gender effects were analyzed with a 2 x 2 analysis of variance (ANOVA) with *type of interaction* and *gender* serving as between-subjects variables and *overall score* as an indicator of game performance. Both main effects were significant, $F(1, 61) \geq 7.12$, $p \leq .01$. More importantly, the interaction barely missed significance, $F(1, 61)=3.94$, $p=.05$. With regard to game success, the advantage for male gamers over female gamers was most prominent when the game is played utilizing a traditional form of game controls (i.e., keyboard and mouse). With regard to utilizing the *Wii-Remote*®, gender differences are less pronounced (Fig. 3).

An additional 2 x 2 ANOVA with *perceived level of game violence* serving as the dependent variable revealed only one significant result; female participants ($M=2.17$, $SD=0.62$) perceived *Wii Remote Dungeon Quest* as significantly more violent[4] than male participants ($M=2.77$, $SD=0.86$), $F(1, 61)=9.62$, $p<.01$, for the main effect (all other $ps \geq .67$). Type of game controls did not affect ratings; participants in the *Gesture* condition ($M=2.43$, $SD=0.67$) and in the *Standard Interaction* condition ($M=2.51$, $SD=0.81$) reported similar levels of perceived game violence.

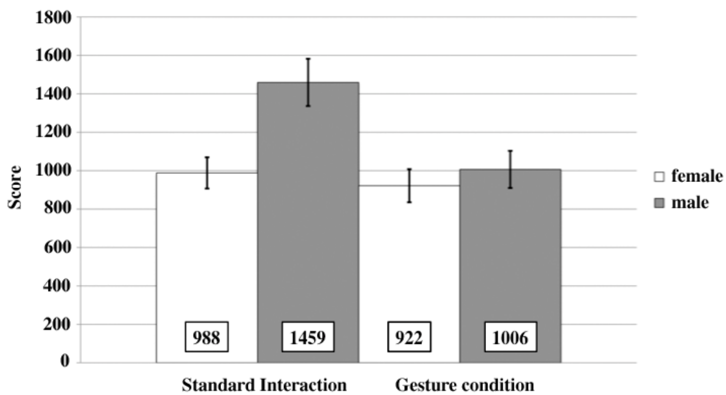


Fig. 3. Mean scores for female and male players of Wii Remote Dungeon Quest as a function of game controls. Error bars indicate standard deviations.

5 Discussion

The present study aimed at analyzing the effects of different forms of videogame controls on behavior, emotion, and cognition. Compared to traditional game controls (e.g., keyboard and mouse), motion detection controls have been suspected of boosting the negative effects reported for violent video games [14]. To test this hypothesis, an action role-playing game was modified that allowed for comparing the effects of different modes of interaction within an otherwise identical game environment. In the game, physical aggression aimed at killing in-game characters was exerted either by mouse clicks or by performing realistic striking movements using motion detection controls (i.e., *Wii-Remote*® control).

---

[4] The item *"how violent do you think this game was?"* was inversely coded with larger numbers indicating lesser degrees of perceived violence.

In contrast to studies that reported greater game success with technologies that also provide high levels of involvement and immersion ([12]), our results clearly indicated greater game success in the *Standard Interaction* condition. With regard to behavioral consequences, thus, traditional forms of playing violent video games that utilize keyboard and mouse appear to be more efficient than realistic yet time-consuming forms of interaction based on performing realistic striking movements.

However, being less efficient in attacking and killing in a violent video game does not at all imply that players will also be less affected otherwise. Rather, when presented with ambiguous conflict scenarios participants in the *Gesture* group showed more hostile cognitions (i.e., anger) than in the *Standard Interaction* condition. The same numerical pattern was found for the other two indicators of aggression (i.e., hostile attribution and desire for revenge).

Performing striking movements is likely to produce greater levels of physiological arousal than pushing mouse buttons, which may result in greater hostility. We know from aggression research that, according to excitation transfer, physiological arousal from previous bodily exertion will intensify subsequent anger reactions in conflict situations, which, in turn, makes aggressive responses of the person more likely [17].

It is important to note that self-reports on emotion clearly indicate that less game success in the *Gesture* group did not entail greater negative affect (e.g., frustration) compared to *Standard Interaction*. Rather, individual game success and the amount of reported negative emotions were significantly *positively* correlated in *both* conditions (r=.26, p=.04); higher scores were thus associated with stronger negative feelings.

Taken together, performing realistic striking movements with motion detection controls in a violent video game might be inefficient in terms of scores, but increases cognitive effects of aggression compared to standard forms of interaction. In addition, players in both conditions were put into bad mood after playing the game. The present study thus is in line with previous findings that playing violent video games increases negative feelings *per se* (e.g., [5]).

Unfortunately, however, we cannot decide whether the emotional effects in the present study result from *game content* or *type of interaction*. Because only a violent game was used, a follow-up study is needed that contrasts the effects of playing a *nonviolent* game (e.g., sport activities) using either motion detection controls or keyboard and mouse. This would clarify whether the increase in negative affect is a general by-product of the greater bodily exertion when performing natural movements (i.e., *type of interaction* effect), or whether it may be specific for violent video games (i.e., effect of *game content*).

The present study also revealed interesting insights with regard to gender differences. We found that the "general male superiority" in game success almost disappeared when participants utilized the *Wii-Remote*® control. This was the case because male players' game performance *dropped* when they utilized the *Wii-Remote*® instead of keyboard and mouse. This finding is likely to reflect differences in learned behavior based on gender specific game preferences; when exposed to a game that allows for transfer of learned forms of game controls, male players are likely to benefit from previous game experiences. There is ample evidence that previous experiences with action games positively transfer to performance in visual tasks (e.g., [18]).

However, when a game requires interaction patterns, which substantially differ from previous experiences such that there is no transfer of learned behavioral patterns (as was the case in the present study with the *Wii-Remote*® control), gender differences disappear.

6 Concluding Remarks

It has been argued that higher levels of involvement and immersion in motion detection controls compared to standard game interfaces affect both psychological factors (i.e., emotion and cognition) and objective parameters (i.e., game success). As to violent video games, this effect is thought to be resulting from performing aggressive activities such as hitting virtual persons utilizing motion detection controls. However, a previous study did not find any indication that motion detection controls affect players more negatively than using traditional forms of game interfaces [14].

In the present study, however, a general increase in negative affect was shown for both groups, and cognitive effects were prominent in the *Gesture* group—especially with regard to anger reactions in conflict situations. Current models of human aggression like GAM [6] hypothesize that interacting internal state variables like arousal, feelings, and thoughts play a major role in triggering real-life aggressive behavior. Based on our findings, thus, giving the all-clear for I/O-devices like the *Wii-Remote*® control would be clearly premature. Rather, future studies will show whether or not the simulation of violent behavior with motion detection controls is indeed psychologically more dangerous than other game interfaces. It will be particularly interesting to test the effects of upcoming entertainment concepts. These include gestural gaming interfaces like Sony's *Move* or Microsoft's controller-free *Kinect* technology ("Project Natal"), which potentially provides even greater experiences of game immersion.

Acknowledgments

The authors wish to thank Florian Hillesheim, Michèle Kallenbach, Stefanie Luxa, Maike Maurer, Petra Pfannkuchen, and Violetta Schaan for running the experiments. We also like to thank the reviewers for their valuable comments and suggestions.

References

1. Laakso, S., Laakso, M.: Design of a body-driven multiplayer game system. Computers in Entertainment (CIE) 4(4), article No. 7 (2006)
2. Ackland, B., Weste, N., Burr, D.J.: An integrated multiprocessing array for time warp pattern matching. In: Proc. 8th Ann. Symp. Computer Architecture, pp. 197–215 (1981)
3. Lee, K.M., Peng, W.: What do we know about social and psychological effects of computer games? A comprehensive review of the current literature. In: Vorderer, P., Bryant, J. (eds.) Playing video games: Motives, responses, and consequences, Mahwah, NJ, LEA, pp. 327–345 (2006)
4. Baron, R.A., Richardson, D.R.: Human aggression. Plenum Press, New York (1994)

5. Anderson, C.A., Shibuya, A., Ihori, N., Swing, E.L., Bushman, B.J., Sakamoto, A., Rothstein, H.R., Saleem, M.: Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. Psych. Bull. 136, 151–173 (2010)
6. Lindsay, J.J., Anderson, C.A.: From antecedent conditions to violent actions: A general affective aggression model. Pers. Soc. Psych. Bull. 26, 533–547 (2000)
7. Anderson, C.A., Carnagey, N.L.: Causal effects of violent sports video games on aggression: Is it competitiveness or violent content? J. Exp. Soc. Psych. 45, 731–739 (2009)
8. Barlett, C.P., Harris, R.J., Bruey, C.: The effect of the amount of blood in a violent video game on aggression, hostility, and arousal. J. Exp. Soc. Psych. 44, 539–546 (2008)
9. Carnagey, N.L., Anderson, C.A.: Violent video game exposure and aggression: A literature review. Minerva Psichiatrica 45, 1–18 (2004)
10. Klimmt, C., Hartmann, T.: Effectance, self-efficacy, and the motivation to play video games. In: Vorderer, P., Bryant, J. (eds.) Playing video games: Motives, responses, and consequences, Mahwah, NJ, LEA, pp. 132–145 (2006)
11. Ivory, J.D., Kalyanaraman, S.: The effects of technological advancement and violent content in video games on players' feelings of presence, involvement, physiological arousal, and aggression. J. Communic. 57, 532–555 (2007)
12. Persky, S., Blascovich, J.: Immersive virtual environments versus traditional platforms: Effects of violent and nonviolent video game play. Media Psych. 10, 135–156 (2007)
13. Huesmann, L.R., Moise, J., Podolski, C.L., Eron, L.: Longitudinal relations between children's exposure to TV violence and their aggressive and violent behavior in young adulthood: 1977–1992. Dev. Psych. 39, 201–221 (2003)
14. Markey, P.M., Scherer, K.: An examination of psychoticism and motion capture controls as moderators of the effects of violent video games. Comp. Hum. Behav. 25, 407–411 (2009)
15. Watson, D., Clark, L.A., Tellegen, A.: Development and validation of brief measures of positive and negative affect: The PANAS scales. J. Pers. Soc. Psych. 54, 1063–1070 (1988)
16. Buss, A.H., Perry, M.: The aggression questionnaire. J. Pers. Soc. Psych. 63, 459–463 (1992)
17. Zillmann, D.: Attribution and misattribution of excitatory reactions. In: Harvey, J.H., Ickes, W.J., Kidd, R.F. (eds.) New directions in attribution research, vol. 2, pp. 335–368. Erlbaum, Hillsdale (1978)
18. Green, C.S., Bavelier, D.: Action-video-game experience alters the spatial resolution of vision. Psych. Sc. 18, 88–94 (2007)
19. Salvador, S., Chan, P.: FastDTW: Toward accurate dynamic time warping in linear time and space. In: Workshop on Mining Temporal and Sequential Data. ACM, New York (2004)

Logos, Pathos, and Entertainment

Ryohei Nakatsu

Interactive & Digital Media Instutite, National University of Singapore
21 Heng Mui Keng Terrace, I-Cube Building Level 2,
Singapore 119613
idmdir@nus.edu.sg

Abstract. Various new forms of entertainment using information and media technologies have emerged and been accepted among people all over the world. Casual and serious games, as well as communication using mobile phones, blogs, and Twitter, are such kinds of new entertainment. It is important to discuss the basic characteristics of such entertainment and to understand the direction to which these new forms are leading human societies. This paper provides a comparative study of entertainment between developing countries and developed countries, and between ancient times and the present day. The future relationship between entertainment and society is also described.

1 Introduction

New network, information, and media technologies are rapidly changing our society, including human relationships, lifestyles, and communication. Entertainment is one area that these new technologies are strongly influencing. One good example is games. Playing computer and video games is a common daily activity for people, especially for younger generations. In particular, online games have become very popular in both the U.S. and Asian countries. Another good example is communication. People communicate with their families and friends through e-mail, mobile phones, texting, Twitter, and other means. Until the 1980s, communication media such as telephones were mainly used for business communication. Today, however, communication extends beyond business conversations and has become a form of everyday entertainment.

There are many discussions on these phenomena, but most of them merely observe what is happening in society and report the fact that more people are spending more time on these new forms of entertainment. Unfortunately, however, there has been little consideration of the basic reasons why these new forms have been accepted by people all over the world. One fundamental questions is whether this a totally new phenomenon within our long history. Another question is which aspects of society are changing through the introduction of these new forms of entertainment, and how much those aspects are changing. Another, more difficult question is the direction to which such entertainment is leading our society. In other words, what will be the future of our society, where people are expected to spend more time on entertainment than they do now.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 183–192, 2010.

In this paper, I try to answer these questions by starting from the question of what entertainment is. It will be clarified that over the long course of human history, people have tried to clearly separate logical and emotional aspects of our behaviors, in other words, the logos and pathos, respectively. Consequently, we have succeeded in confining the great power of emotion to only the private aspects of our lives.

New media such as games and mobile phones, however, have strongly affected this relationship between logos and pathos and have partly destroyed it. Today, people are showing the emotional aspect of their behaviors even in formal situations. A more distinctive point is that this tendency is more obvious in Western countries.

Finally, given these considerations, I anticipate the direction to which our society is headed. I also discuss what we should do to maintain our basic identity as human beings.

2 The Nature of Entertainment

2.1 What Is Entertainment?

At least one billon people currently face starvation all over the world. For these people it is crucial to obtain food for tomorrow or even today. This problem has been one of the most serious topics at global conferences and meetings such as World Economic Forum [1].

On the other hand, in developed countries new types of entertainment have emerged, such as chat on mobile phones and games on game systems and PCs. People in those countries are tending to spend more and more time enjoying such forms of entertainment. There have been significant concerns and complaints about this trend. The basic logic of such complaints is that compared with other human activities, such as education, business, industrial production, and so on, entertainment is not productive. In other words, the complaints suggest that entertainment is only a waste of time.

An important question, however, is whether this is actually true, when the established entertainment industry is huge and includes the movie, game, sports, and other businesses. Another question is why the demand and markets for such entertainment are so huge. The issue of whether entertainment is a waste of time clearly requires care.

2.2 Origin of Entertainment

What is happening now is not a totally new phenomenon. In earlier times, human life was simple. Humans farmed or hunted to survive. When people were not occupied with these tasks, they entertained themselves by various means. In other words, we can say that food is strongly related to our physical sustainability, yet at the same time, we can point out that entertainment is related to our mental and spiritual sustainability.

Then the era of civilization began. People introduced various novel types of activities, such as art, business, learning and teaching, and religion. Because of these activities, entertainment came to be considered as a secondary activity in human life. Although entertainment remained a certain part of our everyday life, it has not been considered an essential part. Figure 1 illustrates these changes in our physical and mental sustainability.

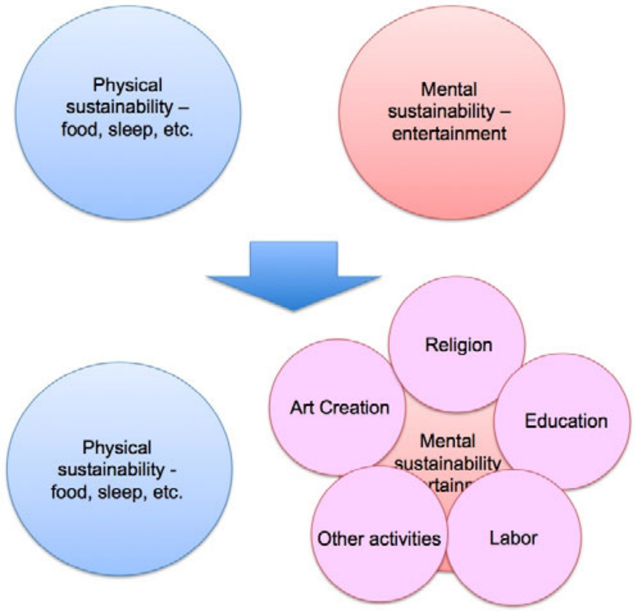


Fig. 1. Changes in the physical and mental sustainability of human beings

2.3 Our Life and Entertainment

Even now sometimes that fact that entertainment is an important part of our life becomes apparent. Consider passengers on an airplane. During a flight, most people sleep, eat, or entertain themselves by watching movies, reading novels, and so forth. Only a few people work during a flight. This means that in a simple situation, our lifestyle consists of three basic activities: sleeping, eating, and entertaining ourselves.

What is happening now is, in one sense, an "entertainment renaissance." The introduction of new technologies, especially interactive technologies, into traditional forms of entertainment has totally renewed and strengthened those forms. People are again noticing the basic strength and meaning of entertainment and recognizing that it is a substantial part of their lives. This is a key point in trying to understand such substantial issues as the role of entertainment in developed and developing countries and the future direction of entertainment.

3 Entertainment in Developing and Developed Countries

3.1 Entertainment in Developing Countries

The importance of entertainment in developing countries has been underestimated. As described above, entertainment has been, from its origin, an essential part of our life. In once sense, its importance is almost second only to survival. Even though few people in developing countries use PCs, mobile phones, or game systems, they intuitively know the importance of entertainment. This is the key point.

Leaders and academics in the developed world have wondered and struggled with how to introduce a higher level of development into developing countries. For example, it was considered difficult to teach the importance of activities such as education, religion, and business to people facing starvation. What is happening in developed countries, however, can simplify this. In developed countries it is observed that an emerging phenomenon of the merger of entertainment with other activities such as education and business. In other words, many activities commonplace in the developed world are becoming forms of entertainment. This is, in one sense, a fundamental change in business model for various human behaviors and activities. Another important point is that this new business model works even without utilizing cutting-edge technologies.

3.2 An Entertainment Renaissance

In other words, the border between entertainment and activities such as education and business is going to disappear. This is why the current phenomenon could be regarded as an "entertainment renaissance."

This means that many serious human activities contain the essence of entertainment and could be enjoyed. This could make it easier to promote such activities as various types of entertainment to people in developing countries. This would lead to greater participation in activities such as education and business, because these activities could now be interpreted as a form of entertainment and would thus be more familiar.

On the other hand, we should carefully examine the new forms of entertainment enthusiastically accepted by younger generations in developed countries. The question is whether these are actually new types of entertainment or not. When we examined new entertainment types in detail, we found that most of them have their origins in older forms of entertainment. For example, the experience of playing role-playing games is almost the same as the experience of reading fantasy novels. Another good example is new types of communication. Communication via chat, texting, or Twitter, using mobile phones, is actually a reshaped form of everyday conversation among family members and friends.

This means that new information and media technologies have reshaped traditional entertainment while keeping the same basic concept. Because of this, while the game market expanded with surprisingly great speed in the early days of video and PC games, it has now apparently reached a certain level of saturation, and various new forms of entertainment face the problem of sustainability. We should probably understand the core concept of entertainment and the future directions of the game industry by studying how people in developing countries entertain themselves.

4 Logos, Pathos, and Entertainment

4.1 Logos and Pathos

What is happening in developed countries is actually deep and substantial. Through the emergence of new forms of entertainment, our societies and lifestyles are experiencing a fundamental change.

Fig. 2. Logos and pathos

Here, we recall the ancient Greek origin of Western philosophy. Plato compared the human spirit to a carriage with two horses and one driver in his *Republic* [2]. Here, as illustrated in Fig. 2, the driver is a metaphor for the rational aspect of the human spirit, called the “logos.” On the other hand, the two horses are a metaphor for the emotional aspect, called the “pathos.” The former could be linked to the formal parts of our lives, and the latter, to the private parts. Furthermore, one of the horses represents passion, while the other represents the instinctive aspect of emotion.

Plato admired logos as the basis of rational human behaviors. As for pathos, passion was admired as the source of creative behaviors, but instinct was despised as undeveloped and representing the dark side of behavior. Plato also expected that logos could control the dark side of pathos, instinct, with help from the bright side of pathos, passion.

This definition and statement by Plato determined the initial direction of philosophy and morality in the Western world. Since the ancient Greek era, people have been trying to separate logos and pathos in their lives. In other words, they have been trying to separate the logical and emotional aspects of living.

4.2 Formal and Informal Aspects of Life

In the long history of Western society, starting with Plato’s philosophical considerations, logos has been considered related to the formal aspects of human behaviors, in social situations and in business scenes, for example. On the other hand, pathos has been related to the private aspects of human activities. Westerners have long and eagerly sought to clearly separate these two aspects. In addition, people have tried to emphasize the importance of logos while neglecting pathos.

Asian languages also make such a distinction. For example, in Japan the formal aspect of behavior is called *honne*, while the informal aspect is called *tatemae*. In terms of *honne* and *tatemae*, Japanese people have been accused of having a double standard of behavior. As the above observation of Western society shows, however, this is not correct. All humans have both formal and private aspects of their behavior, in other words, *honne* and *tatemae*. The problem is that because people in Asian societies have not been very conscious of these two aspects, they have not been good at clearly separating these two types of behavior. In other words, Japanese people could be said to present confusion of the formal and private aspects of behavior.

One good example is that of the former prime minster, Mr. Asoh. He once declared at the Japanese parliament that he was actually against privatization of Japan's Post Office when he was a member of the previous Koizumi cabinet. The problem with his behavior is that, because he was a member of the cabinet and this formal decision was made within the cabinet, it was inappropriate to express his private opinion in a formal situation. What was worse was that he thought he was being honest by expressing his private opinion and could not understand why people accused him of expressing *honne*.

4.3 Approach between Logos and Pathos

In contrast with the above situation, the introduction of new media and forms of entertainment has invaded deeply into our lives and changed our behaviors. It was long considered civilized and sophisticated to act logically and hide the emotional aspect of behavior. Now, however, people tend to expose their emotions even in the formal parts of their lives. A good example is means of communication. It has become common for people to communicate even with close friends or family members by using mobile phones or PCs even during meetings or during dinner. This is surprising because communication with people close to us has typically been a private, emotional behavior.

This means that in our everyday life emotional behaviors have again become influential and can play a major role, after long years of separation between logos and pathos and the higher priority of logos over pathos in formal situations. In one sense, our behaviors are returning to those in ancient times. Moreover, this phenomenon has another important aspect. In observing the phenomenon, we notice two fundamental, distinctive features. First, human behaviors in the Western and Eastern worlds are becoming more similar. Second, human behaviors in developed countries and developing countries are also becoming more similar. In other words, differences in human behavior styles between Western and Asian countries, and between developed and developing countries, are disappearing. People are beginning to share the same ways of thinking and cultural principles and rules. In one sense, this is good because it will lower barriers between different people and countries. At the same time, however, we are losing the local features of cultures, which have long been preserved over the course of history.

5 Media and Entertainment

5.1 History of Media

It is important to consider why we are losing the separation between formal and informal aspects of our behaviors. As indicated above, the arrival of new media strongly influences this phenomenon.

To explain this, consider the history of significant media inventions. The two most impactful media inventions are written characters and the invention of printing typography. These remarkable inventions have enabled people to think, memorize, discuss, and describe by using language. In other words, these inventions have made people more left-brain dependent. It is noteworthy that this mainly happened in Western countries, as shown by the long history of Western philosophy, as represented by Plato. What happened in Asian counties is somewhat different. Somehow people did not try to clearly separate rational and emotional actions. The reason for this is an interesting research topic and is discussed elsewhere. In any case, however, for Asian people the concept of separation between logos and pathos has not been so clear.

5.2 Influence of Movies and Telephones

Two more impactful inventions in the recent history of media are telephones and movies. Today we tend to focus on recent inventions such as video games, mobile phones, e-mail, blogs, and Twitter. Unfortunately, we have almost forgotten the major impact telephones and movies have had on the basic changes in our behavior.

Before the invention of the telephone, the formal and emotional aspects of our behavior were clearly separated. For emotional behaviors, mental distance and special distance were closely related. When we are together with people to whom we are close, such as family members and friends, we expose our private, emotional behaviors. We have a strong instinct to be connected with familiar people. Before the invention of the telephone, however, when we were spatially separated from such people, because we had no method of communicating with them, we had to hide our emotional aspects and behave formally, as in business situations. Then, after the telephone was invented, it enabled us to be connected to familiar people even though we were spatially separated. Since that time, people have gradually tended to mix the formal and emotional aspects of our lives. This is the fundamental reason why now, even during meetings, dinner, or other situations, people want to continue communicating with people to whom they are close by using mobile phones and smart phones. In other words, the telephone has initiated confusion between the rational and emotional aspects of our behavior. The development of mobile phones and other recent media is only accelerating this trend.

The invention of movies has had a similar effect. Before the invention of movies, people were trained to be left-brain dependent, and reading and writing were the major intellectual and communication behaviors. The invention of movies, however, introduced images as an important communication medium. Images have a strong power to directly influence the emotional part of the brain, the right brain. Therefore, after long years of training to rationally use the left brain, people began to depend more on using the right brain. This means that people have gradually become emotionally dependent instead of rationally dependent. The recent trend toward excessive use of computer graphics and animation originated with the invention of movies. These newer technologies are merely accelerating the trend.

5.3 Future of Entertainment and Media

As described above, after long years of separation between logos and pathos in human behavior, these two aspects are moving back together, we can expect that they will merge again in the future, as illustrated in Fig. 3.

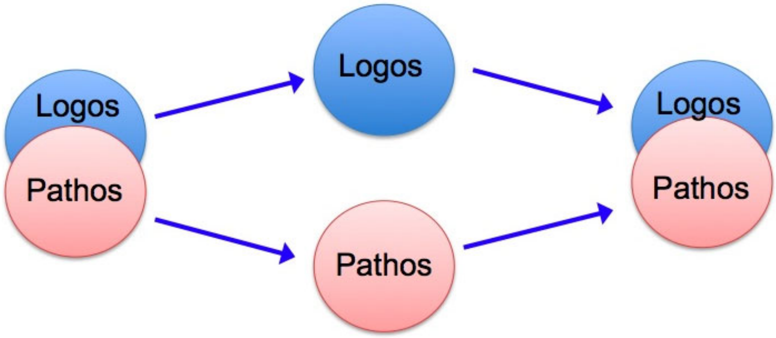


Fig. 3. Transitions of the relationship between logos and pathos

Regarding what this means for the future, it can be pointed out that an interesting and important phenomenon would happen. That is the behaviors of Westerners and Asians could become more alike and even become identical. As described above, in Asian countries a clear separation of logos and pathos has somehow not reached conscious awareness or discussion. Because of this, Asian people have been accused of being underdeveloped as human beings. In Western countries, however, people's behaviors are becoming more like those of Asian people [3]. A good example is communication behaviors. Today, even during such formal events as classes, meetings, discussions, or dinner, people often tend to check e-mail on their mobile phones and try to communicate with people to whom they are close. When a plane lands and arrives at a gate, the first thing many people do is take out their mobile phones, turn them on, and check e-mail or start calling family members. In this regard, there is no distinctive difference between the behaviors of Westerners and Asians.

5.4 Future of Human Beings

In one sense, this phenomenon is good, as various gaps between the West and Asia have long been headaches for cultural understanding. We can probably regard this as a bright side of globalization, but on the other side there is a danger. Each country and each ethnicity has developed its own culture, resulting in a rich variety of cultures all over the world. This is a major human accomplishment. The merger of behaviors between the West and Asia, however, could destroy this rich variety of cultural differences among different countries [4]. It is difficult to anticipate a future in which there are no cultural differences and people all over the world are connected to the network all the time and repeatedly receive and send shallow text messages. It could be like the virtual reality shown in the movie *The Matrix*.

There is another interpretation of the merger of logos and pathos. Animals lack any distinction of logos and pathos; rather, the two are tightly merged. Therefore, the merger of logos and pathos might mean that human beings will revert to an animal-like state. This would another bad outcome, in which we return to a very ancient time when human beings was not yet well developed and our behaviors were almost the same as those of other animals.

Instead, we probably should learn from our history. In both the Western world and the Asian world, there were eras when logos and pathos were not so clearly separated, yet people used to live fully human lives. For example, in the era of Homer's *Iliad* [5], logos and pathos were not clearly separated but the characters behaved honestly, bravely, and heroically. We should compare the behaviors of heroes and heroines in classical antiquity with those of modern people and determine from such comparisons what our behaviors should be in the future.

On the other hand, in Japan we have *The Tale of Genji* [6], the world's oldest novel. In once sense, the behaviors of the heroes and heroines in this novel are somewhat similar to those of current, younger generations. In the classic novel, men and women frequently exchange short poems, called *waka*, as a method of communication. Similarly, young people today frequently exchange short messages or post to Twitter accounts by using mobile phones. Probably the biggest difference, then, between these ancient and modern behaviors, is that the exchange of poems is a form of art creation, whereas message exchange using e-mail and Twitter does not involve art. The major question is how we can ennoble the communication behaviors of people in the network age. A good outcome would be to achieve the way of life described in *The Tale of Genji* but over a network using new media.

6 Conclusion

Various types of new entertainment, such as games and mobile phone applications, have been introduced into our society, rapidly changing our lifestyle. It seems we are in the era of chaos, and it is not certain what kind of lifestyle and society we will have in the future. In this situation, it is important to carefully observe the phenomena happening around us and extract findings, trends, tendencies, and so forth.

In this paper, by trying to answer the question of what entertainment is, I have tried to clarify the basic trends underlying various surface phenomena. It was first clarified that human behavior consists of two aspects, the logical and the emotional, and that entertainment is closely related to the emotional aspect.

Then, I discussed how Plato tried to separate the logical and emotional aspects of human behavior. Plato's efforts originated the long history of Western philosophy. Because entertainment is closely connected to the emotional side of behavior, it has been considered an informal aspect of behavior and thus hidden, even though entertainment is an essential aspect of our lives, like eating and sleeping.

It was also pointed out that the recent development of information and media technologies has raised the importance of the emotional aspect of life. In this sense, we are experiencing an entertainment renaissance. On the other hand, this means that logos and pathos, which have been clearly separated through history, are approaching each other and could even merge together.

There are several dangers in this trend. One is that the cultural differences that have enriched human history might fade away. Another is that human behaviors could revert to those of animals.

So far, there are no clear solutions to overcoming these dangers. One way to address this difficult situation, however, is to look back and learn from our history. There have been several eras when logos and pathos were joined together, as in the times described in the *Iliad* and *The Tale of Genji*. By learning from the behaviors detailed in these famous classics, we could perhaps imagine the future to which we are headed.

References

1. http://www.weforum.org/en/index.htm
2. Ptato: The Rebublic. Penguin Classics
3. Tosa, N.: Cultural Computing. NTT Publishing Co. (2009)
4. Nakatsu, R.: Future of Communications Empowered by Technologies. Ohm-sha publishing Co. (2010)
5. Homer: Ilias. Hackett Publishing Company, Inc., Indianapolis (1997)
6. Murasaki, S.: The Tale of Genji. A Divition of Random House, Inc. (1976)

The Video Cube Puzzle: On Investigating Temporal Coordination

Eric Yim, William Joseph Gaudet, and Sid Fels

Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, B.C., Canada
kyim@ece.ubc.ca, joe@joegaudet.com, ssfels@ece.ubc.ca

Abstract. We have created a novel computer-based 3D puzzle, named *Video Cube Puzzle* to investigate human beings' temporal coordination abilities. Ten adult participants were studied solving ten cubic video puzzles of varying difficulties using a within-subject design. The ten puzzles have two segmentation variations, 2x2x2 and 3x3x3, and five texture variations, solid colours and four videos of drastically different contents. Only 60% of the subjects were able to complete the entire problem set. The results indicate that random imagery and "active" videos make for easier Video Cube Puzzles. Similarly, a geometric increase in difficulty was noted as the number of segments in the puzzle increased. The challenging nature of temporal video cube puzzles appears to be partly due to people's poor ability to process temporal information when using a spatial representation of the timeline using a three dimensional volume. Additional studies are suggested to explore this further. As a new type of game however, the Video Cube Puzzle allows the complexity of the puzzle to be easily varied from simple to extremely complex providing a way to have a continuous pathway to skill and control leading to a satisfying experience when the puzzle is solved.

Keywords: Multimedia, Video Game, Video Cubism, tx-transform, Slit-scan, Puzzle, ePuzzle.

1 Introduction

Jigsaw puzzles have been around since the 18th century. The most common way we see them today is in the form of cut-out cardboard pieces that, when assembled in the correct way, will reproduce a given 2D image. In addition to taking different layouts in 2D, jigsaw puzzles have also taken many shapes and forms throughout its history – they can be categorized into the physical and virtual departments. In the physical department, there are, for example, double-sided jigsaws, puzzles showing optical illusions, 3D jigsaw puzzles that assemble to a castle and various other objects, and a subgroup of 3D jigsaws that are spherical and are also known as puzzle globes. The virtual department, with the help of computer processing, provides even richer variety. For instance, there are virtual puzzles that simply translate the physical ones

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 193–204, 2010.

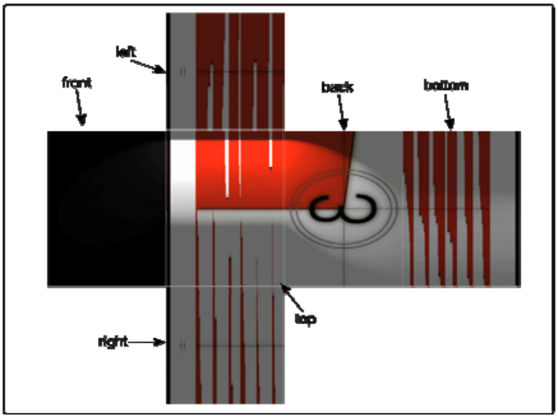


Fig. 1. The faces of a Video Cube, loaded with a countdown video, arranged in paper-model-like fashion. The usually "hidden" sides of the video become a source of interesting visual artifacts exploited by the Video Cube Puzzle.

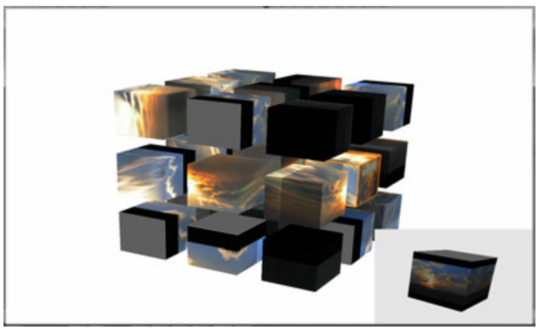


Fig. 2. Interface of the Video Cube Puzzle. The larger view in the middle shows the puzzle portion that the player interacts and solves, whereas the smaller view at the corner shows the reference that the player tries to match.

directly into a virtual environment, popular puzzle video games like Tetris[1] and the Puyo Puyo series[2] which are spin-offs of the central concept, and an audio version that splits up a soundtrack into scrambled segments and requires players to use their ears to solve the puzzle [1]. Despite the diversity in representations, however, the underlying concept has remained grossly unchanged – identify the spatial relationships between puzzle pieces and assemble or rearrange these pieces to achieve a visual goal. In other words, they mainly challenge our spatial coordination.

---

[1] Tetris is a popular puzzle video game that uses falling tetrominoes as its game pieces in a grid. The objectives are to manipulate and stack the falling game pieces to fill and erase as many horizontal lines as possible to collect points and to avoid stacking over the grid.

[2] Puyo Puyo is a popular Japanese falling pieces puzzler. The game is always played against at least one opponent, computer or human. The objective is to outlast your opponent by connecting as many sets of four same-coloured puyos as possible so that they can disappear from your grid and become obstacles on the opponent's grid.

Given the efficiency and effectiveness of the human race in solving such spatial riddles, the objective of this paper is to investigate and discover human beings' abilities to solve a novel virtual puzzle called *Video Cube Puzzle* (VCP), written using OpenSceneGraph (OSG) that has an added temporal dimension of difficulty.

A *Video Cube* is a virtual cube that hosts video frames in its volume rendering it a cubic video player. It is inspired by the work done on Interactive Video Cubism [7, 8]. The front and back faces mark the ends of the video. So, when seen from the front face, the cube appears as a movie screen displaying the first frame of the video. The video can be played on the cube. That is, a parallel, invisible cut-plane passes through the cube from front to back at a constant rate while also stepping through the video frames at the same rate. This way, the entire clip will appear as being played on the cube as it is being consumed. The edges of the video frames form the side images of the cube. These side images are non-existent when the video played in the traditional 2D fashion and, thus, are an interesting artifact of the Video Cube.

An example is shown in Fig. 1. This Video Cube is loaded with a countdown video often seen in films, and its texture-mapped faces are arranged in a paper-model-like fashion for clarity. As clearly seen, all faces, except for the front and back, display strange images that one would not expect to see in a countdown video; yet, these side images are created from the often-ignored edge pixels from the same clip.

The puzzle is formed by segmenting the cube into smaller, equal-sized sub-cubes after which they are randomly displaced in both positions and orientations. Because the video plays only along one direction of the cube, a strange visual effect known as tx-transform [14] or slit-scan [12, 15] can be observed on disoriented sub-cubes. Players are asked to rearrange the sub-cubes to resemble the given solution much like the objective of the Rubik's Cube. But the similarity ends there because, unlike the Rubik's Cube, players must manipulate the cubes individually for this puzzle. The allowable operations are 90-degree rotations around the three principle axes and positions swapping between two cubes. Players are asked to take advantage of the continuum of the video source to solve the puzzle.

In addition to being an interesting artifact, the side images also provide temporal cues that aid in solving the VCP. Although in static form they appear as surface mapped texture that seems to carry only spatial information, these temporal-data-formed images are usually random, irregular, and confusing to a point that would not make sense unless put in a dynamic, temporal context. These temporal cues can effectively convey the subtle dynamic changes of the video clip.

Like traditional jigsaw puzzles, the difficulty of the VCP can be raised by increasing the number of segmentation. Moreover, difficulty can be further raised by choosing a video source with low spatial and/or temporal variations.

2 Related Work

As mentioned previously, there are many varieties in the world of virtual jigsaw puzzles. Many of these computer-based adaptations are considerably easier than their physical counterparts because they do not allow pieces to be rotated, thus reducing the number of degrees of freedom. For example, all the puzzles on the popular website JigZone.com are in that nature. Contrarily, there is a popular kind of on-line jigsaw puzzle, called moving video jigsaw puzzle, that uses a constantly looping video as their image source and, as a result, makes for a more challenging and interesting experience [3]. This is an example of an integration of a video source with a puzzle.

Cube is a common element used in puzzles and games. Some of the interesting applications that use cubes as their main puzzle element include Crazy Cube [2], the Virtual Kanji Puzzle Game [10], and the Cubed Electronic Puzzle Game from ThinkGeek.com [4, 13].

Crazy Cube is an on-line game where players are asked to "[u]se [their] mouse pointer to link every pair of like coloured markers on the cube faces" [2].

The Virtual Kanji Puzzle Game is an augmented reality application that requires the player to put together basic Kanji parts to form a more complicated one. The idea is to help players learn and practice Kanji.

The Cubed Electronic Puzzle Game, unlike the above two, is a physical puzzle game. It "is composed of four electronic cubes. Each one is surrounded on four sides by magnetic connectors, and each displays 1/4 of the puzzle on its LCD screen." [13]. The objective of the game is to connect the four cubes together so that the combined image will form a meaningful shape or a word.

In summary, there are puzzles that use videos as their image source and there are both virtual and physical puzzles and games that take the form of a cube, but there is no puzzles quite like what is proposed. Also, these puzzles and games are used mainly for entertainment purposes only. Although the moving video jigsaw has an obvious temporal component, no academic studies have been done to leverage that feature.

The following examples are neither puzzles nor games, but they are worth mentioning because they are relevant to the subject.

YouCube is a web application that allows users to map YouTube videos of their choice onto the surfaces of an interactive cube. Besides it not being a puzzle, it is different than the Video Cube Puzzle in that it shows multiple videos concurrently and it uses surfaces, not the entire volume, of the cube to show them [11].

Also not a puzzle, but Cubee [17] and other similar devices, such as pCubee [16] and a view-dependent, polyhedral 3D display [9], are cubic fish tank VR displays. The former two are superior in the sense that they are interactive and can act as input devices, as well as displays. They can be interesting displays and more intuitive input devices than the keyboard and mouse for the VCP. And since both of them use OSG to render their graphics, integration ought to be trivial.

Observations and analyses made from this study can potentially inspire design of a new genre of toys and puzzles [5]. Furthermore, insights gained from this study may also inspire creative applications for new displays such as Cubee and the like.

The remainder of this paper is organized in the following manner: Section 3 describes the Video Cube Puzzle and methodology used to conduct the experiment, Section 4 reports results collected from the user study, Section 5 discusses any unapparent findings and observations from the results, Section 6 draws logical conclusions based on the study results, and, finally, Section 7 concludes this article by suggesting future improvements to the proposed application.

3 Study of Human's Effective Use of Temporal Cues

The research question to be studied is simple: *Can human beings effectively utilize temporal cues provided in the form of video to help them solve a novel 3D puzzle?*

3.1 System

The experiments were conducted running the Video Cube Puzzle on either an IBM ThinkPad with hardware accelerated graphics or on a Windows virtual machine running on an Apple Macintosh. The principle methods of human computer interaction were with a mouse and keyboard, the former is responsible for the manipulation of the individual cubes, while the latter is primarily tasked with program control. Fig. 3 depicts a participant engaging in a test session.

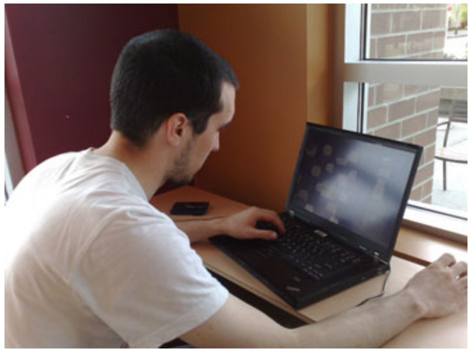


Fig. 3. Participant engaging in a test session

3.1.1 Design

As mentioned previously, the Video Cube Puzzle is written using OSG. The interface is developed with user-friendliness in mind and it is kept to be as clean and uncluttered as possible. The default interface has two views – a puzzle view, which takes up the majority of the screen real estate, and a much smaller solution view on the bottom-right corner of the screen (Fig. 2). Each view has its own event handlers and camera manipulator.

The OSG built-in trackball manipulator is adopted for camera manipulation. Also, users can select, rotate, and swap sub-cubes using the mouse. Fig. 4 shows a selected sub-cube with its rotation handles exposed.

In addition to the mouse inputs, there is a set of keyboard commands that apply to the puzzle. Examples include the ‘e’ key, which toggles the puzzle's expansion feature to expose all faces of the sub-cubes (Fig. 5), and the ‘p’ key, which plays the loaded video on the cube (Fig. 6).

3.2 Experimental Protocol

To mitigate the effects of the “learning curve”, much effort has been made to tune the user interactions via keyboard and mouse for the Video Cube Puzzle. For instance, left-clicking on a sub-cube selects it and exposes its rotation handles; right-clicking on the rotation handle rotates the sub-cube in the indicated direction; right-clicking another sub-cube while one is selected swaps the two; left-clicking-and-dragging rotates the camera around the puzzle; right-clicking-and-dragging pans the camera; and wheel-scrolling zooms the view in and out. Furthermore, tuning of the interface was done at all stages of development so as to continually refine the user experience.

Nevertheless, the user interface requires a certain level of training. As such, before the trials are conducted users are allowed practice “runs” through the system using first a colour only cube, and then with several very trivial cases of the Video Cube, where segmentation variation is held to 2x2x2 and the videos are relatively simple. Users are allowed to repeat such practice runs until they become completely comfortable with the interface.

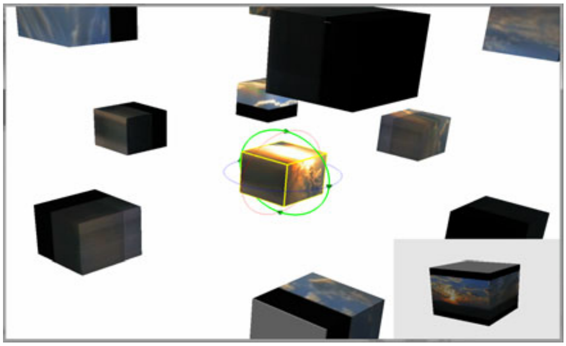


Fig. 4. A screenshot showing a sub-cube being selected, exposing the rotation handles

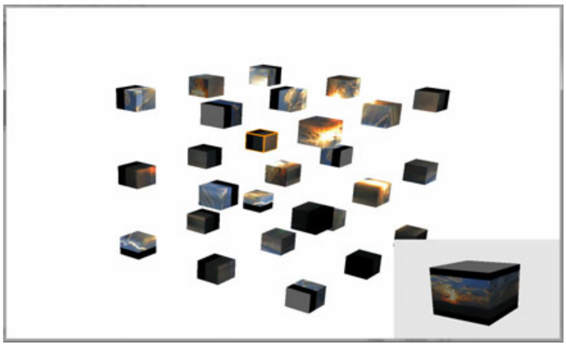


Fig. 5. A screenshot showing the VCP in the expanded state

Fig. 6. A screenshot of a video being played on the VCP

3.2.1 Baseline

A baseline is established to create a common ground where results collected from users of varying puzzle solving experiences can be compared.

The trial procedure starts with solving a simple 2x2x2 and 3x3x3 cubic puzzles with only solid colours for textures (Fig. 7). This allows the development of a percent difference for each trial, whereby a comparison can be made with the baseline colour puzzle and any particular video textured puzzle. This is done using the standard percent difference equation applied to a particular trial as follows:

$$\%\ diff = |t_{baseline} - t_{trial}| \ / \ (t_{baseline} + t_{trial}) \cdot 200 \ . \quad (1)$$

This puzzle uses a simple colouring scheme, with six solid colours (i.e. red, blue, orange, yellow, white and cyan) for its six faces, and a solution is reached when all faces are a continuous single colour, as described in [6].

Fig. 7. The Baseline colour cube puzzles against which each of the trials is compared; from left to right are the 2x2x2, 3x3x3, and the completed puzzle

3.2.2 Surveys

To support the quantitative evaluation outlined in Section 3.2.1, entrance surveys were prepared for each of the candidates in order to assess their previous experience with traditional jigsaw type puzzles, as well as their experience with virtual puzzles. The surveys are designed to determine the user's experience with this type of puzzle, as well as their perception of their own skill at them.

In addition to the entrance surveys, exit surveys are given which have been designed to assess the candidate's feeling with regard to the overall difficulty of the puzzles presented, how they rated the relative increase in the puzzle's difficulty with increased segmentation, and finally how useful the additional temporal element was in solving the puzzle.

3.2.3 Trials

The video were selected to give a broad range of video types from highly regular and predictable to highly random. The goal of selecting these video types was to ascertain the level to which the temporal cues are used as opposed to simply pattern matching the faces of the sub-cubes with the reference cube. Examples of these video puzzles are shown in Fig. 8.

What follows is a description of the trial procedure used to collect the experimental data.

1. Users were given an entrance survey as outlined in Section 3.2.2 to determine their experience with puzzles.
2. Users were given a learning period as described in Section 3.2 to acclimatize themselves with the user interface, and the puzzle in general.
3. Users were given a 2x2x2 and a 3x3x3 colour-only puzzle to establish a baseline for comparison as described in Section 3.2 and shown in Fig. 7.

4. Users were given a puzzle which depicts a setting sun on the horizon this video shows constant but highly predictable change in every pixel, this puzzle was given in 2x2x2 and 3x3x3.
5. Users were given a puzzle which depicts a typical film lead in countdown. This video is highly normalized in so far as most areas of the frame remain constant for large periods of time. This puzzle was given in 2x2x2 and 3x3x3.
6. Users were given a puzzle which depicts a rock band playing in a psychedelic forest. This video is highly irregular showing large amounts of variation over time. This puzzle was given in 2x2x2 and 3x3x3.
7. Users were given a puzzle which depicts a murder of crows flying against a grey and highly regular sky. This video creates a non-discernible pattern on all sides of the Video Cube, both spatially and temporally. This puzzle was given in 2x2x2 and 3x3x3.
8. Users were given an exit survey to determine some qualitative measures about their interaction with the Video Cube Puzzle, and the efficacy of temporal information.

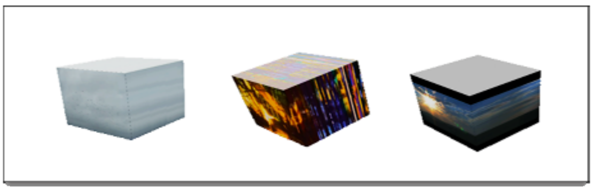


Fig. 8. Examples of the Video Cube Puzzles used

4 Results

Ten trials were conducted on a broad range of subjects with varying experience with puzzle solving. A chart of results of the average performances is shown in Fig. 9.

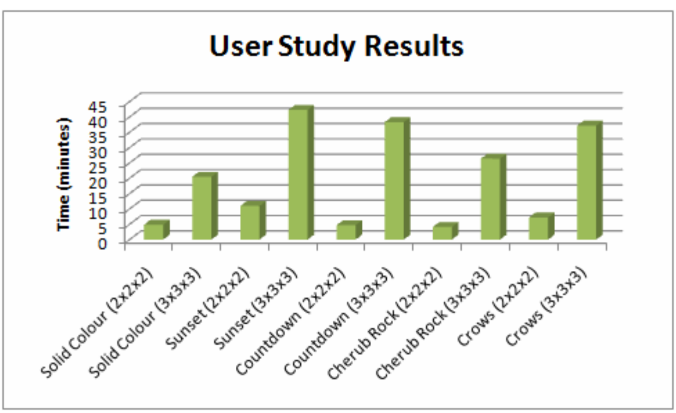


Fig. 9. This chart shows the average times, in minutes, subjects took to complete the puzzles

4.1 Surveys

Based on our entrance surveys collected from the participants, the average perceived efficacy with standard jigsaw type puzzles was found to be 7.25 (1 being terrible and 10 being very effective), with a standard deviation of 0.71. Furthermore the average proficiency with virtual puzzles (i.e. computer-based puzzles) was 7.625, with a standard deviation of 0.75.

The results of the exit surveys, which pertained specifically to the difficulty of the puzzle, are as follows:

1. Compared to a traditional jigsaw puzzle, on a scale of 1 to 10 rank the relative difficulty of the Video Cube Puzzle (1 being trivial, 10 being extremely difficult). The average response of the candidates was a 7, with a standard deviation of 0.8.
2. If the 2x2x2 Video Cube Puzzle has a difficulty of 1, how would you rate the difficulty of the 3x3x3 (1 being the same, 10 being extremely difficult)? The mean response was 7.4 with a standard deviation of 1.78.
3. How useful did you find the temporal cues given by the video (1 being useless, 10 being critical)? The mean usefulness was seen to be 6.1 with a standard deviation of 2.18.

5 Discussion

While the number of trials conducted seems small, it should be noted that the mean time for completion of the full problem set was 2.9 hours. Furthermore, several candidates found the video puzzle prohibitively difficult to the point of failing to complete the entire problem set. The total mean completion time for a study, finished or unfinished, was 2.42 hours with a standard deviation of 1.06 hours.

Using the formula outlined in Section 3.2.1, relative percent difference calculations were made for each of the trials. A chart of these results is shown in Fig. 10.

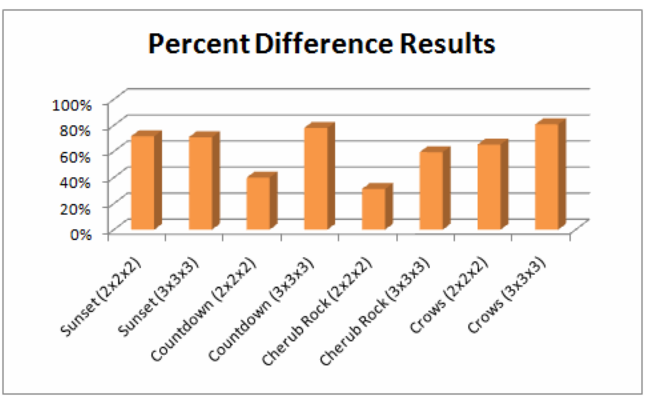


Fig. 10. This chart shows the percent differences of the averages of video cube results and their corresponding averaged baseline solid-colour results

While the surveyed results show that subjects seemed on average to think the temporal cues were somewhat helpful, observations made by the test givers suggest they seemed more inclined to treat the puzzle as sub-jigsaw puzzles. Furthermore, subjects would often fail to realize the continuum from one sub-cube to another, that is to say two consecutive cubes would have similar textures on their opposing faces. This continuity can be leveraged to more easily solve the puzzle.

The poor performances may also be attributed to an unaccustomed user interface. Two out of ten subjects interviewed expressed frustration with the UI complaining that it did not operate according to conventions established by the CAD tools they were accustomed. Refinement and modification are to be made in a future version to address this issue.

A couple of interesting observations are made from the results: first of all, majority of the participants found the first video-textured cube, namely the Sunset, extremely hard, and then many found the Cherub Rock much easier than the rest. For the Sunset, its results may partly be attributed to it being the first encounter of temporal data and partly to the fact that it is a, both spatially and temporally, slowly varying video; several participants complained about the clouds in the video looking the same across frames. This assertion is further reinforced by the results of the Cherub Rock, which is the video with the most activities and best overall performance. Additional tests using similarly “active” videos will verify this assertion.

6 Conclusions

The results from the experiment have made any conclusive assertions difficult. First and foremost, the fact that 40% of subjects were incapable of completing the puzzle, and those that did take on average 3.05 hours to complete, seems to indicate that, in the general case, randomly selected humans were incapable of leveraging the temporal data provided to assist them in the solving of the puzzle. For a frame of reference, the developers, on average, took 1.66 hours to complete the problem set; however, they had a high degree of familiarity with both the nature of the puzzle and the mechanics of the interface.

The videos chosen were an attempt to ascertain which type of video data would provide the most assistance, and conversely which video data would make the puzzle the most difficult. The Sunset puzzle was seen to be highly temporal, where each pixel in the movie changed in a highly predictable manner, the Cherub Rock video was seen as entirely random thus providing little to no useful temporal data, and the Crows video was seen as containing minimal amounts of temporal data. Based on the results, it appears that the more random and “active” videos provided less of a challenge to the users than the more linear and “predictably” changing videos. Similarly, it was noted that there was a geometric increase in difficulty as the number of segments in the puzzle increased.

Due to the limited sample set, it is at this point not possible to make a definitive assertion about the overall ability of humans to process temporal information when presented in a three dimensional volume. It is the hope of the authors that a future study with a greater sample set could be conducted to obtain more definitive results.

7 Future Work

7.1 Interface

Exit interviews with the candidates resulted in a common feature request – the ability to “lock" elements of the puzzle once they had been situated. In addition, the controls for rotating individual sub-cubes were said to be too fine, and prone to either mis-clicks, or double-clicks. Refinements on both of these controls and other re-mapping suggested by test participants are required to get more accurate results from users.

7.2 Puzzle Design

The authors have assumed, without grounds, cubic puzzle pieces are appropriate for this study. This fact is to be verified in future studies. Testing using several different geometries, such as pyramids and octagonal prisms, is necessary to gain scientific insights.

Additionally, as discussed in Section 5 testing using more videos, particularly highly varying ones, is required to verify the assertion made about which types of videos would make easier/harder Video Cube Puzzles.

References

1. Audio puzzle, `http://www.audiopuzzle.com`
2. Crazy cube, `http://www.epuzzles.com/crazy-cube/index.htm`
3. Jigsaw puzzles free, `http://www.jigsaw-puzzles-free.com/VideoJigsaws/Video-Jigsaw-Puzzles.aspx`
4. Thinkgeek, `http://www.thinkgeek.com`
5. Barr, P., Noble, J., Biddle, R.: Video Game Values: Human-Computer Interaction and Games. Interact. Comput. 19(2), 180–195 (2007)
6. Berkove, J., Sickle, J., Hummon, B., Kogut, J.: An Analysis of the (colored cubes)^3 puzzle. Discrete Mathematics 308(7), 1033–1045 (2008)
7. Fels, S., Lee, E., Mase, K.: Techniques for Interactive Video Cubism (poster session). In: 8th ACM International Conference on Multimedia, New York, pp. 368–370 (2000)
8. Fels, S., Mase, K.: Interactive Video Cubism. In: 1999 Workshop on New Paradigms in Information Visualization and Manipulation in conjunction with the 8th ACM International Conference on Information and Knowledge Management, New York, pp. 78–82 (1999)
9. Harish, P., Narayanan, P.J.: A View-dependent, Polyhedral 3D Display. In: 8th International Conference on Virtual Reality Continuum and its Applications in Industry, New York, pp. 71–75 (2009)
10. Hirose, J., Guo, Q., Yamamoto, T., Hirayama, M.J.: A Virtual Kanji Puzzle Game Based on 3D Graphics and an Intuitive Inputting Device. In: ACM SIGGRAPH ASIA 2009 Sketches, New York, p. 1 (2009)
11. Meyers, A.: YouCube, `http://www.universaloscillation.com/youcube/`
12. Parke, F.I.: Adaptation of Scan and Slit-scan Techniques to Computer Animation. SIGGRAPH Comput. Graph. 14(3), 178–181 (1980)

13. Piriloo, C.: What's Your Favorite Puzzle Game of All Time?, `http://chris.pirillo.com/whats-your-favorite-puzzle-game-of-all-time/`
14. Reinhart, M.: Tx-transform, `http://www.tx-transform.com/Eng/index.html`
15. Sims, L.: Animation, Scan, and Slit Scan Photography. In: 25th Intl. Tech. Comm. Conf., Dallas (1978)
16. Stavness, I., Lam, B., Fels, S.: pCubee: A Perspective-corrected Handheld Cubic Display. In: 28th International Conference on Human Factors in Computing Systems, New York, pp. 1381–1390 (2010)
17. Stavness, I., Vogt, F., Fels, S.: Cubee: A Cubic 3D Display for Physics-based Interaction. In: ACM SIGGRAPH 2006 Sketches, New York, p. 165 (2006)

Emotions: The Voice of the Unconscious

Matthias Rauterberg

Industrial Design, Eindhoven University of Technology, Den Dolech 2
5600MB Eindhoven, The Netherlands
g.w.m.rauterberg@tue.nl

Abstract. In the paper the idea is presented that emotions are the result of a high dimensional optimization process happening in the unconscious mapped onto the low dimensional conscious. Instead of framing emotions as a separate subcomponent of our cognitive architecture, we argue for emotions as the main characteristic of the communication between the unconscious and the conscious. We see emotions as the conscious experiences of affect based on complex internal states. Based on this holistic view we recommend a different design and architecture for entertainment robots and other entertainment products with 'emotional' behavior. Intuition is the powerful information processing function of the unconscious while emotion is the result of this process communicated to the conscious. Emotions are the perception of the mapping from the high dimensional problem solving space of the unconscious to the low dimensional space of the conscious.

Keywords: emotion, intuition, unconscious, conscious, mapping.

1 Introduction

A lot of concepts and frameworks about emotions are already available [1-2]. A good definition of "emotion" is a notorious problem. Depending on the conceptualization and operationalization of the phenomenon to be studied, different kind of research is not sufficiently progressing towards consensus. These contributions challenge researchers in the behavioral and cognitive sciences in the importance of definitions and their consequences for distinguishing related but fundamentally different affective processes, states, and traits.

In this paper we argue for a new holistic view of the relation between the unconscious and the conscious information processing part of our brain. Although a lot of research has been done since the 80s of last century, still important questions are unanswered [3]. The most accepted view on emotion is that it is a modular subcomponent of our cognitive system linked and related to a lot of cognitive functions [4]. This conceptualization is also leading the design and architecture of entertainment robots [5]. We will argue for a different view so that designers of such kind of systems are better supported than nowadays.

Several of the main characteristics of emotions are their richness, heterogeneity, vagueness and openness for multiple interpretations [6]. This is one of the most interesting but also often overlooked and underestimated aspect of emotions. We will argue for a new way of how to conceptualize emotions as a central aspect between the unconscious and the conscious information processing.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 205–215, 2010.

2 Emotion, Unconscious and the Mapping Problem

In this section throughout we discuss the state of art and beyond for concepts about emotion, unconscious and mappings from high dimensional spaces to low dimensional spaces.

2.1 Concepts of Emotions: The State of Art

Several cognitive functions can be ordered according to their required internal processing time throughout life span (from milliseconds for reflexes till months for moods and drives; see Figure 1). Proper design of entertainment systems has the potential to stimulate and influence most of these functions. The primary cognitive functions are: *reflexes*, *sensations*, *thoughts*, *dreams*, *emotions*, *moods*, and *drives*.

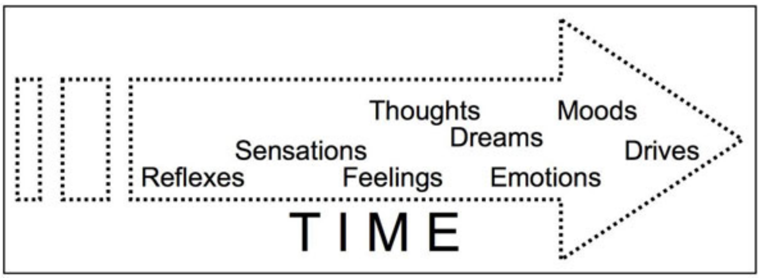


Fig. 1. Time scale of some cognitive functions (adopted from [7])

These different cognitive functions are linked to different control systems (bold black in Figure 2) of our body. In turn, these links help us design the right interaction (italic in Figure 2) through various body parts and control systems. To achieve users' emotional involvement, one needs to address these interactions with the right channels (as the examples given to the right of Figure 2). "Emotional states are defined by myriad changes in the body's chemical profile; by changes in the state of viscera; and by changes in the degree of concentration of varied striated muscles of the face, throat, trunk, and limbs. But they are also defined by changes in the collection of neural structures which cause those changes to occur in the first place and which also cause other significant changes in the state of several neural circuits within the brain itself" [2, p. 282]. The human unconscious can be framed by the genetic reproduction system, the peripheral nervous system (PNS) and the central nervous system (CNS). These three distinct systems contribute to unconscious information processing. Each of them has certain subsystems which are directly related to emotional feelings (see Figure 2). One of the main differences between these subsystems is the cycle time to process incoming signals to actions executed (from milliseconds to hours or even months; see Figure 1). All of these processes are related in one or the other way to affects and emotions [8].

Research in psychology has tried to structure prototypical emotional feelings into discrete basic categories [9-10]. Because emotions are complex, diverse and with multiple facets there are different ways to construct these basic categories. The boundaries of the phenomenon *emotion* are so blurry that almost every feeling, mood and other internal perceptions can be categorized as an emotion. The phenomenon *emotion* is too broad to fit into one single scientific category. Emotions vary along certain dimensions (i.e. intensity, amount of pleasure, degree of activation, etc.).

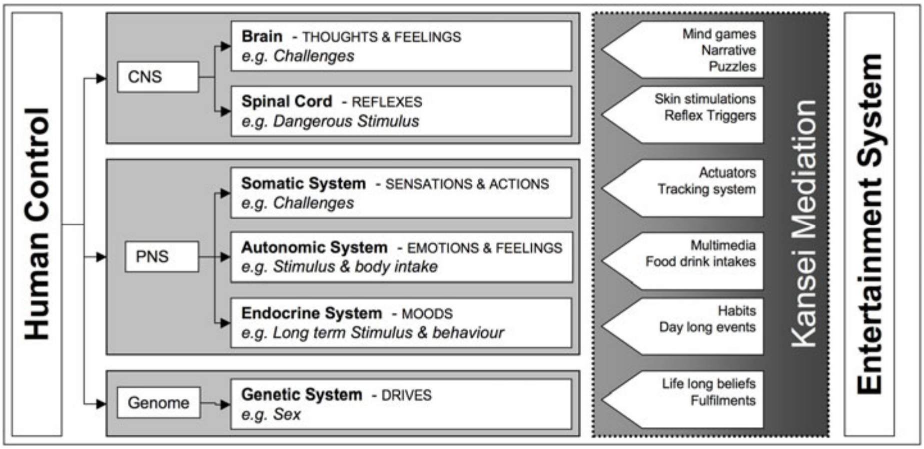


Fig. 2. From human control mechanisms to entertainment [Central Nervous System (CNS), Peripheral Nervous System (PNS)] (adopted from [7])

One of the most prominent models is based on two dimensions: 'unpleasant-pleasant' and 'activation-deactivation' as the core affect [11]. "Emotion categories do not cluster at the axes, and thus the structure of emotion has been said to be a circumplex. Nevertheless, although we are among those who emphasize such findings, we now believe this dimensional structure represents and is limited to the core affect involved. Prototypical emotional episodes fall into only certain regions of the circumplex" [9, p. 807] (see Figure 3).

The combination of the two object-less dimensions 'pleasure-displeasure' and 'activation-deactivation' might capture most emotions but certainly not all. Cognitive information processing, intuition and behavioral planning can account for the myriad manifestations of emotions. The approach so far to capture emotions is dimensional and in addition limited. "The process of changing core affect is not fully understood, but the important point here is the complexity of the causal story" [12, p. 148].

Lane [13] goes even a step further by relating neural correlates to conscious emotional experience. He put forward a hierarchical model as follows [neuroanatomical structure/psychological function; however a one-to-one mapping between neuroanatomy and psychology is not intended]: brainstem/visceral activation, diencephalon/action tendencies, limbic/discrete emotion, paralimbic/blends of emotion, and prefrontal cortex/blends of blends. We can conclude that emotions can be based on conscious and unconscious information processing. Both processes have influences on actual behavior, behavior control and internal adaptation through learning.

Although emotions are complex phenomena and rich in content, main stream psychology tries to capture them in less complex models and frameworks. If we want to maintain the richness in our understanding of emotions, we probably have to change our view. Before we come back to this we will first introduce the unconscious cognitive functions.

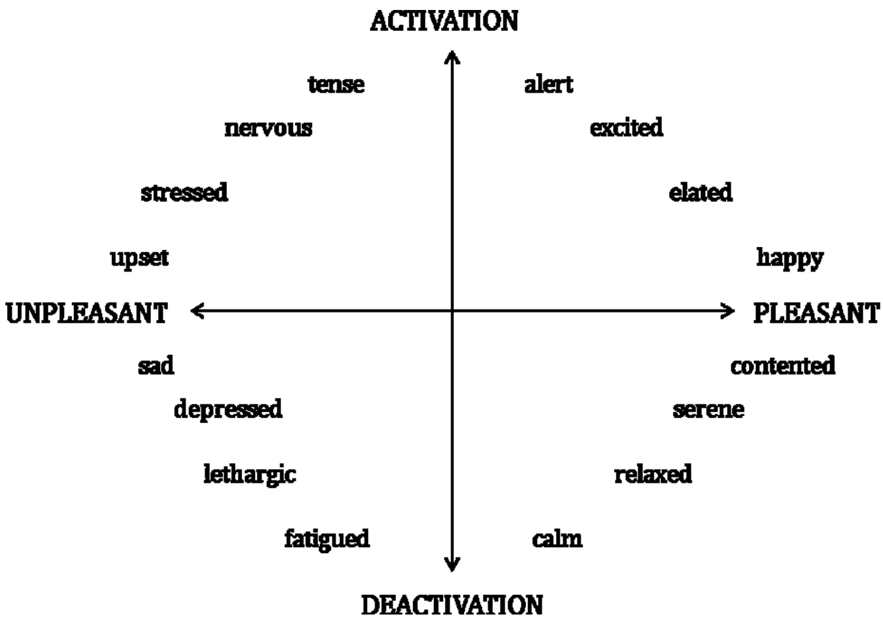


Fig. 3. The two dimensional core affect model (adapted from [12])

2.2 The Power of the Unconscious

When we think of being conscious, we think of being awake and aware of our surroundings (see [14] for more). Being conscious also means being aware of ourselves as individuals. Mostly, people tend to think of being conscious as being alive. We tend to think that the person should be responsive to the surrounding environment to be conscious. Being in a coma is considered to be the opposite of conscious, so called non-conscious or unconsciousness. There are at least three forms of consciousness for humans: (1) the conscious state; (2) the subconscious state; and (3) the unconscious state. In the scope of this paper the unconscious state is fully operational and functional for a normal human living as a parallel background process of our mind and body, we are just not aware of, e.g. activities of the cerebellum. The subconscious can be turned into conscious (i.e. by paying attention to subconscious activities); the unconscious normally is not available to the conscious. The remaining question is how – if at all- does the unconscious communicate with the conscious?

The conscious part of the brain is investigated already for a long time. One of the important results is the limitations of the information processing capacity of the short-term memory [15]. In his classic paper Miller [16] found that the conscious information processing capacity is limited to seven (plus or minus two) chunks or dimensions [17]. This conscious part is mainly described as the short-term or working memory to emphasize its role in decision making and controlling behavior.

Consciousness is a topic for which either there exist no acceptable description, definition and explanation or, and this depends on one's point of view, there are far too many and far too divergent ones. Most definitions from the Western world are resulting in a dualism [18]. This dualism has created a schism between mind and body that does not necessarily exist and that has been a key, not necessarily a correct one, in the Western world understanding of consciousness. Even today's literature is full of reference to the mind and the body as if it has been established beyond doubt that there was indeed a separation [19]. The illusory Cartesian self is more and more challenged by biological and neurofunctional evidences that point to consciousness as an emergent property of competing and successive brain processes. Unconscious and conscious play a role in the initiation and performance of voluntary actions [20-22]. Both action and expression originate in the unconscious [23]. They are probably then vetoed by emotions and moods [24]. Actions and expressions not vetoed are then performed [25].

In the dualist approach, the mind is conscious and the body unconscious. Outside cognitive and brain sciences was and still is a primacy of conscious over unconscious [1]. Interestingly enough, recently one of the pioneers of artificial intelligence started to recognize the importance of the sub-, respectively the unconscious [26]. The unconscious activities of the human mind are hidden under and are controlled by consciousness (the word oppressed is often used). In the emergent view however, there is no such separation between mind and body and consciousness is said to be an emerging property of unconsciousness. It is moving away from the concept of conscious oppression of perception and expression into the concept of emerging perception and expression. Nakatsu, Rauterberg and Salem [22] show a model of the different views of the relationship between human consciousness and sub-consciousness, now (dualist view) and in the future (emergent view).

An iceberg can serve as a useful metaphor to understand the unconscious mind [27-28]. As an iceberg floats in the water, the huge mass of it remains below the surface. Only a small percentage of the whole iceberg is visible above the surface. In this way, the iceberg is like the mind. The conscious mind is what we notice above the surface while the sub- and unconscious mind, the largest and most powerful part, remains unseen below the surface. The unconscious mind holds all awareness that is not presently in the conscious mind. All activities, memories and thoughts that are out of conscious awareness are by definition sub- or even unconscious. Scherer [28] assumes that a large part of emotions functions in an unconscious mode and only some parts will emerge into conscious. But what does this relationship between unconscious and conscious look like?

Recent investigation in cognitive psychology of the conscious and unconscious are promising. Most of the brain's energy consumption is *not* used for processing responses to external stimuli as usually assumed; but what is this enormous amount of brain energy then for? [29-30] One promising aspect of unconscious information processing is finding optimal solutions in the multidimensional sensor and knowledge space of the unconscious for controlling behavior by situated forecasting throughout the near and far future. But how does the unconscious communicate these 'solutions' to the conscious? Before we can provide an answer, we have to introduce the 'mapping problem' from a high (i.e. unconscious) to a low dimensional (i.e. conscious) processing space. The following chapter is purely metaphorical, all introduces concepts have no technical meaning in the context of this paper.

2.3 Mapping from High to Low Dimensional Spaces

A standard problem in several research areas is the visualization of results found in a high dimensional space into the two dimensional (2D) space of a plane surface. This problem is quite old and several solutions are already developed, depending on the particular mapping problem [31-33]. To introduce into these approaches we will shortly discuss this classic 3D-2D example. Abbott [34] wrote the book 'Flatland'. Flatland has only two dimensions, and is populated with lines (females), triangles, squares, polygons, and circles; these inhabitants perceive their 2D environment differently than we do perceive our three dimensional (3D) environment. The basic idea is to explain the main differences about a fourth dimension beyond our 3D world. Flatlanders cannot understand a third dimension, and we have the same trouble with 4D, and even for higher dimensions. In Flatland for example is the ball from a 3D world perceived as changing sizes of circles, the 2D 'shadow' (see Figure 4).

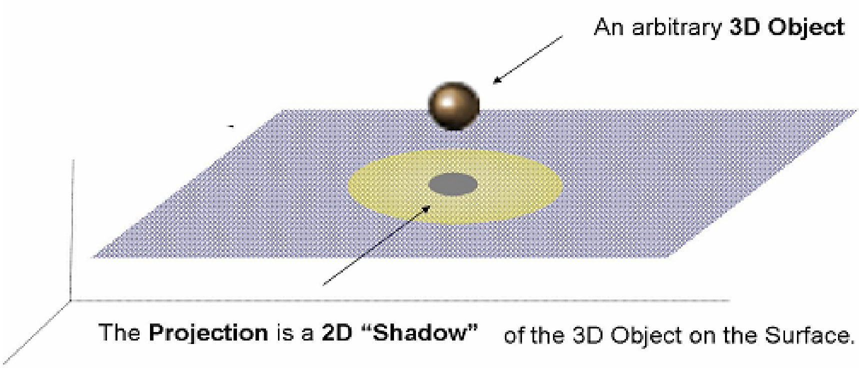


Fig. 4. The projection of a ball shaped 3D object on a 2D surface results in its form of a flat circle (from http://mdchristian.com/Normal_Projections.html)

Rucker [35] goes beyond Abbott's idea by taking the 2D world of the Flatlanders into curved space, black holes, and beyond. Banchoff [36] describes the problem of mappings between different dimensional spaces in more technical terms, applying several mapping methods to the design of computer graphics and other graphical shapes. He uses techniques of slices, projections, shadows, and generalization. The most practical part is learning to count the number of faces, vertices, and edges in a 4D (and higher) hypercube. The mapping downwards into a three dimensional space of a hypercube results in eight different cubes. If we now assume that the unconscious information processing often described as intuition [37-38] takes place in a high dimensional space, then we have to question how the solutions found in this high dimensional space can be mapped into a low dimensional space, the conscious reasoning?

3 From Unconscious to Conscious

Decades ago, Dreyfus questioned already the rational approach of cognitivism by excluding intuition, etc. [39]. Since then luckily a lot of research – in particular in psychology –have shown the growing interests in phenomena like tacit knowledge [40], intuition [37-38] and the unconscious [41]. Scherer assumes that a large majority of cognitive processes related to emotions are unconscious and "that only some of these processes (or their outcomes) will emerge into consciousness for some time" [28, p. 312]. First we have to show that the conscious and the unconscious are separated but related. Kahneman clearly operates on the assumption that both systems are distinct [42], and the unconscious is fallible. He describes the unconscious as 'intuition' and the conscious as 'reasoning'; both systems have clear complementary characteristics (see Table 1). Dienes and Scott conclude that the structural knowledge is really divided between the unconscious and the conscious [43].

Table 1. The two primary cognitive systems: intuition/unconscious and reasoning/conscious. (Adapted from [42, p. 1451])

INTUITION	REASONING
Fast processing	Slow processing
Parallel processing	Serial processing
Automatic processing	Conscious controlled
Effortless	Effortful
Associative	Rule governed
Slow learning	Flexible, adaptive
Emotional / 'hot'	Neutral / 'cold'

Recent investigations in cognitive psychology of the nature of conscious and unconscious information processing are promising. Most of the brain's energy consumption is *not* used for processing responses to external stimuli; but what is this brain energy then for? [29-30]. One possible answer is for sub- or even unconscious information processing that guides behavior through situated forcasting. "Emotions are part of the biological solution to the problem of how to plan and carry out action aimed at satisfying multiple goals in environments which are not perfectly predictable" [11, p. 35]. According to the 'deliberation-without-attention' effect [44] it is not always advantageous to engage in intensive conscious 'decision making' alone. On the basis of recent insights into the characteristics of conscious and unconscious thought, Dijksterhuis et al. [27] tested the hypothesis that simple choices produce better results after conscious information processing, but that choices in complex situations should be left to unconscious information processing. It was confirmed in several studies on consumer choice that purchases of complex products were viewed more favorably when decisions had been made in the absence of attentive deliberation. In addition, Dijksterhuis, Bos, Nordgren, and van Baaren [30, 45] could show the advantages of the unconscious information processing for complex processing, and therefore they open a door to a new view on the relationship towards the power of the unconscious [22, 44, 46]. Dienes and Scott showed in their experiments that conscious structural knowledge is associated with greater consistency in making errors than the unconscious [43, p. 348].

4 Discussion and Conclusions

The emotion experience and other cognitive activities are not separate, independent and distinctive processes, but should be conceptualized as a gradient in the interaction between cognitive activities (i.e. thoughts, memories, beliefs, etc.) [6]. "Brain structures at the heart of neural circuitry for emotion (e.g., the amygdale) impact cognitive processing from early attention allocation through perceptual processing to memory" [6, p. 390]. It seems not possible to explain how neural activities instantiates emotions if we conceptualize emotions as an independent cognitive process.

If we assume that emotions are perceived as important aspects in relation with other cognitive functions than we could go so far to conceptualize emotions as the appearance of these cognitive processes to our conscious. This is an internal perception loop about the own mental and bodily states [11, 47]. If we assume further that the information processing capacity of the unconscious is several magnitudes higher than the conscious, and both systems are somehow separate systems, we have to answer the question how do these two systems communicate with each other. Our idea is that emotions can play this role as the 'voice of the unconscious' in telling the conscious the solutions found in a high dimensional space. Anecdotal evidence shows that excellent ideas (i.e. Eureka experiences) are accompanied by very positive emotions. This is a way of our unconscious telling us something great is happening [48]. We see emotions as the conscious experiences of affect based on unconscious complex internal states and processes. This unconscious processing ends in an emotion as the result of a high dimensional optimization process to be further processed by the conscious part. But these emotions are not only to inform the conscious, they also communicate to the social context around us. Our whole body language is also part of the emotional expression space for the adjustment of social relations [11, 49].

Applying this view to the design of entertainment systems, in particular entertainment robots with human like behavior, we recommend implementing a high dimensional processing unit (mainly sensor data input related to models of the systems itself as of the environment) that maps the found solution of situated forecasting into a low dimensional action control unit, instead of implementing a separate emotion unit [4-5]. This mapping is 'colored' as emotions, primarily for external purposes in social communication expressed via the nonverbal behavior of the entertainment system. In this sense non-verbal expressions are communicating directly to the unconscious of the 'social' environment.

5 Future Work

To realize the conclusions above in entertainment or other interactive products we have to change our whole view on how to design such kind of systems. Although a lot of bio inspired software and hardware architectures are already part of the state of art, we recommend to focus first on the split between unconscious (UPU) versus conscious processing units (CPU) ('intuition' versus 'reasoning' as described in Table 1).

The UPU is mainly characterized by fast, parallel processing based on associative principles [50-51] with an associative memory structure [52]. The main functions of this UPU are problem solving in a high dimensional space and the situational forecasting [14]. Due to the time required for sensor input, signal processing and actuator output the system is always time wise behind reality. To bridge this (even small) time lack forecasting based on situational models of the environment (incl. most important events and objects in it) is needed (similar to [53]). For the CPU standard software and hardware architectures are probably sufficient. The communication between UPU and CPU has to be expressed on the system's surface as a kind of non-verbal expressions with emotional values [54]. The main remaining question is how to specify and implement the communication between UPU and CPU, the voice of the unconscious.

Acknowledgements. About 25 years ago I was heavily engaged in a discussion with Rainer Landmann in Hamburg (Germany) during a long night in front of an open fire. Thanks to this friend I can present the main idea in this paper. As you can see, most of the references are younger than 25 years; hence this is the main reason why I had to wait so long. More recently, I am also very grateful for the constructive comments of Ap Dijksterhuis on earlier versions of this paper.

References

1. Ortony, A., Clore, G.L., Collins, A.: The cognitive structure of emotions. Cambridge University Press, Cambridge (1988)
2. Damasio, A.: The feeling of what happens - Body, emotion and the making of consciousness. Vintage, London (2000)
3. Ortony, A., Turner, T.J.: What's basic about basic emotions? Psychological Review 97, 315–331 (1990)
4. Dörner, D., Hille, K.: Artificial souls: Motivated emotional robots. In: IEEE International Conference on Systems, Man and Cybernetics: Intelligent Systems for the 21st Century, pp. 3828–3832. IEEE Press, Piscataway (1995)
5. Itoh, K., et al.: Mechanisms and functions for a humanoid robot to express human-like emotions. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 4390–4392. IEEE Press, Los Alamitos (2006)
6. Feldman-Barrett, L., et al.: The experience of emotion. Annual Review of Psychology 58, 373–403 (2007)
7. Salem, B., Nakatsu, R., Rauterberg, M.: Kansei mediated entertainment. In: Harper, R., Rauterberg, M., Combetto, M. (eds.) ICEC 2006. LNCS, vol. 4161, pp. 103–116. Springer, Heidelberg (2006)
8. Rolls, E.T.: Precis of the brain and emotion. Behavioral and Brain Sciences 23(2), 177–234 (2000)
9. Russell, J.A., Barrett, L.F.: Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. Journal of Personality and Social Psychology 76(5), 805–819 (1999)
10. O'Rorke, P., Ortony, A.: Explaining emotions. Cognitive Science 18, 283–329 (1994)
11. Oatley, K., Johnson-Laird, P.N.: Towards a cognitive theory of emotions. Cognition and Emotion 1(1), 29–50 (1987)
12. Russell, J.A.: Core affect and the psychological construct of emotion. Psychological Review 110(1), 145–172 (2003)

13. Lane, R.D.: Neural Correlates of Conscious Emotional Experience. In: Lane, R.D.(ed.) Series in Affective Science, pp. 345–370. Oxford University Press, Oxford (2000)
14. Rauterberg, M.: Hypercomputation, unconsciousness and entertainment technology. In: Markopoulos, P., et al. (eds.) Fun and Games 2008, pp. 11–20. Springer, Heidelberg (2008)
15. Baddeley, A.: The magical number seven: Still magic after all these years? Psychological Review 101(2), 353–356 (1994)
16. Miller, G.A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. The Psychological Review 63(2), 81–97 (1956)
17. Simon, H.A.: How big is a chunk?: By combining data from several experiments, a basic human memory unit can be identified and measured. Science 183(4124), 482–488 (1974)
18. Boyle, G.J.: Critical review of state-trait curiosity test development. Motivation and Emotion 7(4), 377–397 (1983)
19. Dreyfus, H.L., Dreyfus, S.E.: From Socrates to expert systems: The limits of calculative rationality. Technology in Society 6(3), 217–233 (1984)
20. Kim, J.: Mind in a physical world - An essay on the mind-body problem and mental causation. MIT Press, Boston (2000)
21. Dijksterhuis, A., Aarts, H.: Goals, attention, and (un)consciousness. Annual Review of Psychology 61(1), 467–490 (2010)
22. Nakatsu, R., Rauterberg, M., Salem, B.: Forms and theories of communication: From multimedia to Kansei mediation. Multimedia Systems 11(3), 304–312 (2006)
23. Libet, B.: Can conscious experience affect brain activity? Journal of Consciousness Studies 10(12), 24–28 (2003)
24. Brass, M., Haggard, P.: To do or not to do: The neural signature of self-control. The Journal of Neuroscience 27(34), 9141–9145 (2007)
25. Libet, B.: Do we have free will? Journal of Consciousness Studies 6(8-9), 47–57 (1999)
26. Schank, R.C.: The computer's subconscious. Applied Artificial Intelligence: An International Journal 23(3), 186–203 (2009)
27. Dijksterhuis, A., Nordgren, L.: A theory of unconscious thought. Perspectives on Psychology 1(2), 95–109 (2006)
28. Scherer, K.R.: Unconscious processes in emotion. In: Feldman-Barrett, L., Niedenthal, P.M., Winkielman, P. (eds.) Emotion and Consciousness, pp. 312–334. Guildford, New York (2005)
29. Raichle, M.E.: The brain's dark energy. Science 314, 1249–1250 (2006)
30. Raichle, M.E.: A brief history of human brain mapping. Trends in Neurosciences 32(2), 118–126 (2008)
31. Tufte, E.R.: Envisioning information. Graphics Press, Cheshire (1990)
32. Tufte, E.R.: The visual display of quantitative information. Graphics Press, Cheshire (1983)
33. Tufte, E.R.: Visual explanations. Graphics Press, Cheshire (1997)
34. Abbott, E.A.: Flatland: A romance of many dimensions. Dover Publications, Mineola (1992)
35. Rucker, R.: The fourth dimension: A guided tour of the higher universe. Houghton Mifflin, Boston (1984)
36. Banchoff, T.F.: Beyond the third dimension: Geometry, computer graphics, and higher dimensions. W.H. Freeman & Company, New York (1996)
37. Khatri, N., Ng, H.A.: The role of intuition in strategic decision making. Human Relations 53(1), 57–86 (2000)

38. Witteman, C., et al.: Assessing rational and intuitive thinking styles. European Journal of Psychological Assessment 25(1), 39–47 (2009)
39. Dreyfus, H.L.: The Socratic and Platonic basis of cognitivism. AI & Society 2, 99–112 (1988)
40. Reber, A.S.: Implicit learning and tacit knowledge. Journal of Experimental Psychology: General 118(3), 219–235 (1989)
41. Naccache, L., Dehaene, S.: Unconscious semantic priming extends to novel unseen stimuli. Cognition 80, 215–229 (2001)
42. Kahneman, D.: Maps of bounded rationality: Psychology for behavioral economics. The American Economic Review 93(5), 1449–1475 (2003)
43. Dienes, Z., Scott, R.: Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. Psychological Research 69, 338–351 (2005)
44. Dijksterhuis, A., et al.: On making the right choice: the deliberation-without-attention effect. Science 311, 1005–1007 (2006)
45. Coutrix, C., Nigay, L.: Mixed reality: A model of mixed interaction. In: Proceedings of International Working Conference on Advanced Visual Interface - AVI 2006, pp. 43–50. ACM, New York (2006)
46. Dehaene, S., et al.: Imaging unconscious semantic priming. Nature 395, 597–600 (1998)
47. Rauterberg, M.: About a framework for information and information processing of learning systems. In: Falkenberg, E.D., Hesse, W., Olivé, A. (eds.) Information System Concepts: Towards a Consolidation of Views, pp. 54–69. Chapman & Hall, London (1995)
48. Dijksterhuis, A.: Emotion and the unconsciousness, Personal Communication with Matthias Rauterberg (2010)
49. Barakova, E., Laurens, T.: Expressing and interpreting emotional movements in social games with robots. In: Personal and Ubiquitous Computing (2010) (online January 16, 2010)
50. Krikelis, A., Weems, C.C.: Associative processing and processors. Computer 27(11), 12–17 (1994)
51. Tan, A.-H.: Adaptive resonance associative map. Neural Networks 8(3), 437–446 (1995)
52. Zeng, Z., Wang, J.: Associative memories based on continuous-time cellular neural networks designed using space-invariant cloning templates. Neural Networks 22(5-6), 651–657 (2009)
53. Wolski, R., Spring, N., Peterson, C.: Implementing a performance forecasting system for metacomputing: the Network Weather Service. In: Proceedings of the 1997 ACM/IEEE Conference on Supercomputing (CDROM), pp. 1–19. ACM, San Jose (1997)
54. Nakatsu, R., Rauterberg, M., Vorderer, P.: A new framework for entertainment computing: From passive to active experience. In: Kishino, F., Kitamura, Y., Kato, H., Nagata, N. (eds.) ICEC 2005. LNCS, vol. 3711, pp. 1–12. Springer, Heidelberg (2005)

Analyzing the Parameters of Prey-Predator Models for Simulation Games

Seongdong Kim[1], Christoph Hoffmann[2], and Varun Ramachandran[2]

[1] Dept. Gameware, Kaywon School of Art & Design
Euiwhang, Republic of Korea 437-080
Tel.: 31-420-1881; Fax: 31-420-1879
sdkim@kaywon.ac.kr
[2] Dept. Computer Science, Purdue University
West Lafayette, USA 47907
Tel.: 765-494-6185; Fax: 765-494-0739
{cmh,vramacha}@purdue.edu

Abstract. We describe and analyze emergent behavior and its effect for a class of prey-predators' simulation models. The simulation uses rule-based agent behavior and follows a prey-predator structure modulated by a number of user-assigned parameters. As part of our analysis, we present key parameter estimations for mapping the prey-predators' simulation parameters to a functional relationship with the LV(Lotka-Volterra) model, and how the parameters interact and drive the evolution of the simulation.

Keywords: prey-predator systems, Lotka-Volterra model, agent-based simulation, emergent behavior, simulation games, parameter estimation.

1 Introduction

Simulation games have become a widely accepted and needed tool for training in government agencies and the military. For example, commercial simulation games have had a dramatic effect on the military's education and training programs [1]. Particularly, in computer games, whether entertainment or educational games, NPCs are agents whose behavior is scripted and automated. In this paper, we develop and test a framework for estimating the impact of specific behavior traits on the overall evolution of game worlds. We restrict to the important subclass of prey-predator simulations that have applications in ecology, biology, economics and sociology, and for which mathematical concepts can be formulated that summarize system properties. For instance, [2] reports a discrete simulation tool to simulate the behavior of predator programs and computer viruses in small scale (1600 node) computer networks. Multi-behavioral games between predators and prey have been studied that integrated pre-encounter and post-encounter behaviors [6]. Mitchell and Lima [8] investigated the shell game using an individual-based model that allows predators to update information about prey location. Huige [9] presented results to predict that very little overlap will occur between the players' waiting distributions and that the

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 216–223, 2010.

predator will rarely outwait the prey. We have applied game theory differently, with prey searching for grazing and predators hunting prey in packs. We show that food availability and predators' densities influence patterns of prey distribution. We approach the problem by relating the (extended) Lotka-Volterra (LV) equations [3], [7] to expose essential characteristics of prey-predators simulations. Illustrating our concepts, we relate quantitative parameters, such as birth rate, life span, time to starvation and so on, to the mathematical model, and so try to clarify how changing the driving parameters impacts the simulation.

2 Relationship to the Lotka-Volterra Model

2.1 Simulation Specifications

Our basic simulation program has been described in [3]. It uses the *OpenSteeer* library [4] to implement basic motion algorithms. The parameter set is as in [3]. For example, using subscript 0 to indicate that the parameter applies to prey, and subscript 1 to indicate that it applies to predators we have:

- *Birth, aging and death:* The L_k is the maximum life span; no agent lives longer than that. Moreover, f_k is the adulthood fraction, that is, the fraction of L_k spent as an adult. The concept here is that juvenile agents have fewer capabilities, qualitatively or quantitatively, than adults. In particular, juveniles do not have offspring as yet. The A_k is the current age, normalized to be between 0 and 1 as the fraction *current age* : *max lifetime*. The h_k is the hunger rate, measuring the speed with which an agent becomes hungry again, and H is the hunger value. The agent dies if $A_k \geq 1$ or if $A_k < 1$ and $r \in \left[0, {A_k}^{18}\right]$. Here r is a random number between 0 and 1, with uniform distribution. Death from hunger depends on H. Initially, if $H = 0$ then hunger is incremented by

$$\Delta H = h_k * \Delta t * \min\left(1, \frac{A_k}{1-f_k}\right) \; . \tag{1}$$

  A death due to hunger occurs if $H > 0$ and $r < H^{10}$ where r is a random number in $[0,1]$. Feeding diminishes the hunger.

2.2 The Lotka-Volterra Model

The LV differential equations express the global behavior of a prey-predator system under the assumption of unlimited food supply for prey. Excluding the trivial steady state in which one or both population types are extinct, the steady state is a dynamic balance of the two populations and is given to a cyclic boom-bust cycle. The equations are

$$X' = \alpha X - \beta XY \; . \tag{2}$$

$$Y' = -\gamma Y + \delta XY . \tag{3}$$

where X and Y are the populations of prey and predators, respectively. The factor α represents the growth rate of the prey population absent predation and so accounts for the birth rate as well as for the rate of death from old age; β is the impact of predation, accounting for death of prey due to being killed by predators. The assumption of unlimited food supply is manifest in that α only assumes death from natural causes. The factor γ quantifies the rate of predators' death from natural causes; δ measures the growth rate of the predator population due to the realized food supply, i.e., due to hunting. Again, change rates are assumed to depend linearly on the population sizes. Assuming nonzero population sizes, steady state is achieved by a zero change rate; that is, when

$$\alpha = \beta Y \ . \tag{4}$$

$$\gamma = \delta X \ . \tag{5}$$

See also [5]. The factors α and γ measure intrinsic characteristics of prey and predators: prey life span and fertility determine α, and predator life span only determines γ. The presence of X in the term δXY implies that predators' fertility depends on the food supply X, an assumption justified in many situations. The simulation architecture assumes a limited food supply for prey, so the equations have to be modified accordingly. We assume a stable prey population size X_c, absent predation, that is due to the carrying capacity of the environment. Once reached, birth and natural death rates of prey should become equal, without predation. Assuming linearity, we modify the LV equations to incorporate these considerations:

$$X' = \alpha\left(1 - \frac{X}{X_c}\right)X - \beta XY. \tag{6}$$

$$Y' = \ -\gamma Y + \delta X. \tag{7}$$

If we assume no predators, then the first equation becomes the logistics equation:

$$X' = \alpha\left(1 - \frac{X}{X_c}\right)X \ . \tag{8}$$

Assuming the continuous case, the solution is:

$$X(t) = X_c\, X(0)/\big(X(0) + \big(X_c - X(0)\big)e^{-\alpha t}\big) \ . \tag{9}$$

Here $X(t)$ is the population at time *t*.

2.3 Carrying Capacity Determination and Base Configuration

Beginning with the logistic equation, we note that the ratio of population size to carrying capacity, for $\alpha > 0$, is 1 in the limit. If $\alpha < 0$, then the ratio goes to zero. The assumption $\alpha > 0$ is the case where the population persists and reaches the carrying capacity of the environment, while for negative α the prey becomes extinct. In the following, we assume that $\alpha > 0$. Note that the magnitude of α only affects how fast the system reaches the carrying capacity. In the stochastic case, we expect that the population at first grows past the carrying capacity and subsequently collapses due to environmental exhaustion.

2.4 Determination of α, β, γ, δ

To determine α, we can run the simulation in the base configuration without any predators. Then the parameter α tracks the average difference of the number of births and deaths at each time step. That is, the intrinsic rate of prey population increase can be estimated as $\alpha(t) = (b(t) - d(t))/X(t)$, where $b(t)$ is the number of births and $d(t)$ is the number of deaths (due to hunger or old age) of the prey, at time *t*. We can use a moving average to compute this estimate. Since the environment has limited food supply, however, the estimate needs to take into account the logistical equation. The formula is then

$$\alpha(t) = \frac{(b(t)-d(t))}{(X_c-X(t))}\frac{X_c}{X(t)} \ . \tag{10}$$

When $X(t)$ approaches X_c the denominator vanishes and the coefficient $\alpha(t)$ cannot be determined. Thus, we have to estimate α by measuring population growth with infinite food supply. To accomplish this, we set the food growth parameters to instantaneous regeneration and set the hunger value to zero. The result is an exponential increase in population size where we determine the value of α by fitting the exponential $X(t) = X(0) \cdot (1 + \alpha)^t$. This should be done for several initial population sizes, arriving at an average value for α.

The factor β measures the impact of predation on the prey population. The equations assume linearity in both predators' and prey densities. We determine the factor using the procedure of Sharov [7]: we run the simulation with one predator for a number of time steps during which no new predators are born.[1] The k-values [7] of β for a single time step are averaged over the simulation interval $[t_0, t_1]$. With *N* the number of steps we obtain

$$-\ln(1 - \beta(t)) = \sum_t \left(-\ln\left(1 - \frac{k(t)}{X(t)}\right)\right)/N \ . \tag{11}$$

At start-up, the simulation requires a number of steps to reach a semi-stable state. Therefore, we have programmed the possibility of running the simulation for a predetermined number of steps before introducing the predators at a random location. The parameters γ and δ measure the attrition of the predator population by natural causes, that is, death from old age and hunger, and the effect on the population when prey is available. The parameters are estimated as follows:

1) For fixed population size X_0 run the simulation for some time, holding the prey population size constant.
2) For X_0, estimate the aggregate predator growth rate *r* from $X_1(t) = X_1(0)(1 + r)^t$.

[1] We assume asexual reproduction for simplicity of the model.

Here, $X_1(t)$ is the predator population size at time t. This is done for a number of different population sizes X_0. Having determined r, for different prey population sizes, the parameters for the predator population dynamics is found by fitting a line to the data provided by the (r, X_0) data:

$$r = \delta X(t) - \gamma \; . \tag{12}$$

2.5 Fitness Metrics

Intuitively, a growth in β indicates improving hunting skills of the predators, while a decrease of β indicates that prey is better able to evade predators and defend themselves against predation. We can therefore use β as a competitive fitness metric by which to quantify the survivability of the two agent populations. Similarly, the other parameters can be considered as indirectly contributing to the competitiveness: Larger α implies greater resilience under predation as well as greater impact on the environment; an increase in $|\gamma|$ implies greater mortality of predators, and larger δ indicates a greater dependence on prey.

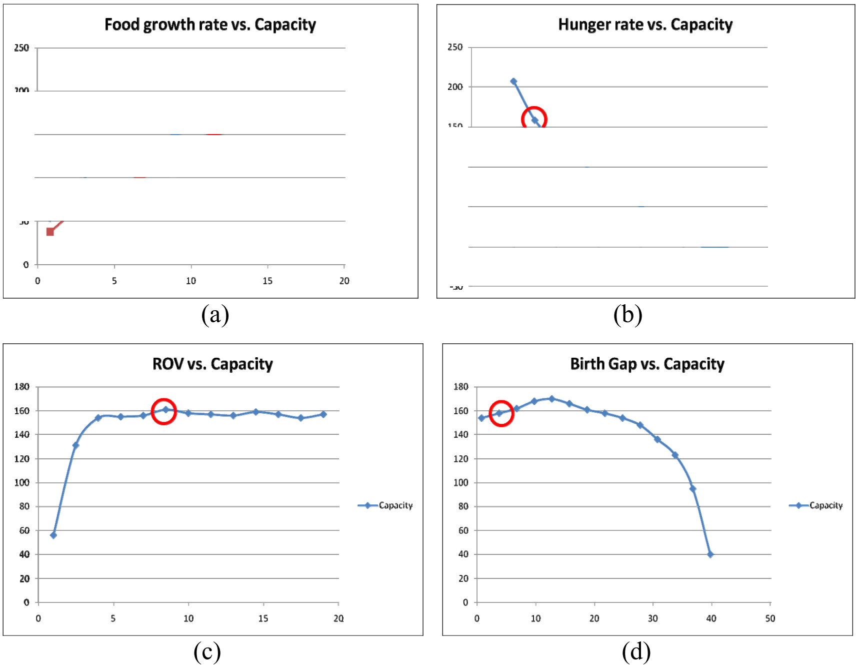
3 Experimental Results

3.1 Parameter Determination

We determine how the many simulation parameters impact the prey-predator system by estimating the coefficients of the mathematical model of Section 2. This provides guidance when seeking to fine-tune a simulation scenario. In our model, each food square in the environment, upon depletion, lies fallow for a time determined by the *food-growth-delay* and then vegetation grows back at a speed set by the *food-growth-rate* parameter. These parameters govern the amount of consumable food per time unit, and directly impact the carrying capacity. Fig.1a shows the correlation fixing all other parameters. The base configuration values are indicated in red.

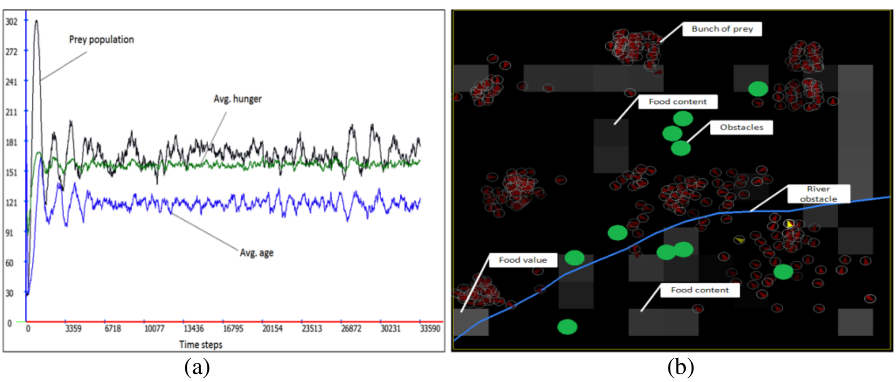
The graphs are relatively straightforward. For instance, the carrying capacity first goes up almost linearly with the food growth rate. However, as the food-growth rate increases, the carrying capacity does not continue to increase linearly and instead approaches a limit value induced by the food-growth-delay. The reason for this is as follows: Beyond a certain rate, re-growth is essentially instantaneous. But grazing takes time and then the food patch has to be fallow for the time specified by the food-growth-delay parameter. This means that the total amount of consumable food cannot exceed a maximum imposed by the rate at which prey grazes and the delay before the food values is fully restored. Fig. 1(a) and (b) show the dependence of carrying capacity on the re-growth speed and regeneration delay. For the base configuration the carrying capacity is about 158. Fig. 2 (a) shows the various measured quantities throughout the simulation of the base configuration.

Fig. 2(b) shows that the simulation of prey-predators' problem implements the model described above. We initialized each simulation with 40 prey and 10 predators and ran the simulation until both the prey and predators' populations became extinct. The square represents the world used in our experiments. The grey patches represent



(a) (b) (c) (d)

Fig. 1. Effect of environmental settings on the carrying capacity (a). Prey characteristics and their effect (b-d). The red circle marks the base configuration values.



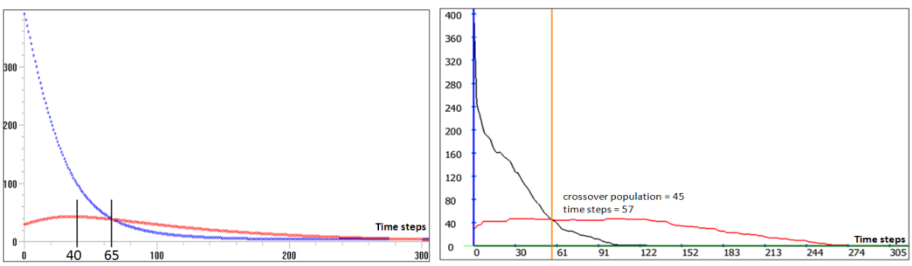
(a) (b)

Fig. 2. The prey population (black curve) and average hunger value (green) of prey. The average age (blue) shows large variation initially accounting for significant fluctuations in (a). The models of simulation game; Red triangles represent live prey while yellow triangles indicate dead prey (b).

food content for the prey. The brightness of a patch indicates its food value. As prey consume a patch, the food value slowly fades out and only grows back after a certain period of time. The agents are represented by the triangular shapes with the red ones prey and the turquoise and pink ones predators. When a prey dies it remains on the screen for some time, turning yellow and slowly fading out. This does not happen for the predators.

3.2 Estimation of the LV Parameters

In Section 2.1 we determined the carrying capacity in the base configuration from a number of simulation runs, without predators, obtaining $X_c \approx 158$. Following the method described in Sharov[7] the four key parameters, in the base configuration, are about $\alpha = 0.075$, $\beta = 0.00543$, $\gamma = 0.0689$, and $\delta = 0.0031$. We integrated the modified LV equations, for a more intense world, with $\alpha = 0.21$, $\beta = 0.024$, $\gamma = 0.23$, and $\delta = 0.002$. These values are the result of more concentrated food growth and hungrier predators. The predicted system evolution is shown in Fig.3 (a):



(a) (b)

Fig. 3. Predicted evolution: The first peak of predator population (red curve) is 42 after about 40 time steps; the crossover of the predator and prey populations (black curve) is 40 after an x-value of about 65 time steps. Carrying capacity is 703 (a). Observed evolution: Predator population peaks at 46, holding steady between time steps 25 and 120; crossover is at time step 57 with 45 individuals (b).

The predictions are in reasonable agreement with the observation, accounting for stochasticity of the simulation. The observed times of extinction are reasonably close considering that, unlike in the equation model, we deal with integer values in the simulation. We have run additional observations and predictions.

4 Discussion and Conclusions

We have investigated and estimated experimentally four key parameters from experiments with the prey-predator simulations that allow predicting the fluctuations of the populations using the augmented LV equation. The configuration space expressing the system evolution as function of the parameter assignments is a high-dimensional manifold, so local exploration around a particular configuration is an expedient way to assess the survival value resulting from the various parameter settings when making adjustments.

Our simulation differs from many other studies by basing prey behavior on Reynolds's flocking algorithms [3]. It is known that the parameter settings of the three basic rules of flocking impact the emergent behavior of a flock [10]. This can influence herd size and cohesion and, with them, the effectiveness of grazing: a large herd could exhaust a sector of the environment so quickly, that not all herd members would have adequate grazing. So, as the herd moves to a new square, the followers in the herd could conceivably starve. Thus, it might be better for prey to roam the environment in herds that are smaller and split and explore in different directions searching for food. Such geometric considerations can be analyzed using our approach.

Acknowledgements

This work has been partially supported by NSF Grants CPATH 0722210 and 0938999, by DOE award DE-FG52-06NA26290, and by a gift from Intel Corporation and by the National Research Foundation of Korea Grant funded by the Korean Government (MEST) (NRF-2010-0016003, Development of Simulation Framework of Group Behaviors for Heterogeneous-Autonomous Character in Computer Game.

References

1. Macedonia, M.: Games, Simulation, and the Military Education Dilemma. In: Forum for the future of Higher Education. Forterra System Inc. (2002)
2. Gupta, A., DuVarney, D.C.: Using Predators to Combat Worms and Viruses: A Simulation-Based Study. In: 20th Annual Computer Security Applications Conference (ACSAC 2004), pp. 116–125 (2004)
3. Kim, S., Hoffmann, C., Lee, J.M.: An Experiment in Rule-based Crowd Behavior for Intelligent Games. In: Fourth, ICCIT (2009), `http://www.cs.purdue.edu/homes/cmh/distribution/PapersChron/3896a410.pdf`
4. Reynolds, C.: The OpenSteer Library (2004), `http://opensteer.sourceforge.net/`
5. Lotka-Volterra equation, Wikipedia (2004)-9
6. Mitchell, W.A.: Multi-behavioral strategies in a predator-prey game: an evolutionary algorithm analysis. Journal comp. Oikos 118, 1073–1083 (2009)
7. Sharov, A.: Quantitative population ecology, `http://home.comcast.net/~sharov/PopEcol/`
8. Mitchell, W.A., Lima, S.L.: Predator-prey shell games: large-scale movement and its implications for decision-making by prey. Journal compilation, Oikos 99, 249–259 (2002)
9. Hugie, D.M.: The waiting game: a battle of waits between predator and prey. International Society for Behavior Ecology 14(6), 807–817 (2003)
10. Chiang, S.-C., Hoffmann, C., Mittal, S.: Emergent Crowd Behavior. In: CAD 2009, Reno, NV (June 2009)

Analyzing Computer Game Narratives

Clark Verbrugge and Peng Zhang*

School of Computer Science, McGill University
Montréal, Québec, Canada, H3A 2A7
{clump,pzhang13}@cs.mcgill.ca

Abstract. In many computer games narrative is a core component with the game centering on an unfolding, interactive storyline which both motivates and is driven by the game-play. Analyzing narratives to ensure good properties is thus important, but scalability remains a barrier to practical use. Here we develop a formal analysis system for interactive fiction narratives. Our approach is based on a relatively high-level game language, and borrows analysis techniques from compiler optimization to improve performance. We demonstrate our system on a variety of non-trivial narratives analyzing a basic reachability problem, the path to win the game. We are able to analyze narratives orders of magnitude larger than the previous state-of-the-art based on lower-level representations. This level of performance allows for verification of narrative properties at practical scales.

Keywords: game narrative,verification,optimization,performance analysis.

1 Introduction

Narrative is a central feature of many computer games, extending from text-based interactive fiction to current adventure, RPG, and mixed genres. For such games a number of narrative properties become important to an immersive game experience, including logical consistency, continuity of story elements [1], level of "tension" or atmosphere [2], as well as game-play issues such as ensuring player progress, and adequate coverage of potential player choices [3,4]. The complexity of larger narratives, however, can make formal analysis of narratives difficult. For analysis of long game segments or existing, complete narratives the combinatorial explosion implied by a rich set of player choices and game content elements results in scaling issues, limiting conservative verification of narrative properties.

In this work we develop an efficient system for analyzing complex game narratives. Our approach is based on an exhaustive analysis of the narrative state-space, and thus applies to general reachability problems. To improve scalability, we extract high-level game information from game code by applying program analysis techniques more traditionally found in the compiler optimization domain. This information is used within an optimized search to reduce the branching factor and improve performance.

We illustrate our design by examining a fundamental example of reachability, the path to win the game. In this context we analyze a variety of non-trivial narratives

* Now at Microsoft: pezhang@microsoft.com.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 224–231, 2010.

derived from both commercial and amateur interactive fiction games. Using our optimized state search we are able to determine winnability within seconds to minutes on a modern machine. These results dramatically improve on previous, low-level work on formal winnability analysis, which has been limited to narrative inputs orders of magnitude smaller [5]. The ability to perform reachability analyses on narratives of much larger scales shows that formal verificaton of a wide variety of properties can be feasibly performed on non-trivial, industrial-size game narratives.

Contributions of our work include the design of a complete system for reachability analysis of computer game narratives, an optimized search based on a novel dataflow analysis, and non-trivial experimental investigation of basic winnability. In the sections that follow we give related work and contextual background for our research, followed by our system design and results.

2 Related Work

Our approach to game narratives aims at analysis/verification in order to improve design and thus player experience. A number of other works have also identified and formalized narrative properties that players find pleasing, disruptive, or which may improve story generation. Adams, for instance, proposed basic guidelines for industry developers in a series of online articles [3], where among many other properties the problem of narrative game-play being allowed past the point of winnability was presented; Verbrugge formalized this notion as "pointlessness." [4]. Barros and Musse proposed a model designed to ensure pace/tension [2], and Nelson et al. defined properties related to the spatial locality of actions, as well as motivational or logical continuity of events [1]. Logic-based approaches are popular in this context, and have been applied to verify the time-line of a non-linear story [6], and ensure logical correctness and continuity of events [7]. Petri net models have also been proposed and used to represent narratives for both story construction and analysis [4,8,9,10]. Our work here builds on previous experience in winnability analysis, applying a generic SAT-based solver to a Petri net model of game narratives [5].

In this work we *post-facto* analyze narratives manually developed as text-based, interactive fiction adventures. Although text-based IF games are an older genre, they have had as a whole a strong influence on more modern interactive fiction, adventure, and RPG game design. An active community continues to exist as well, and a number of IF authoring kits are available as products on their own [11,12] or in terms of extensions built on older commercial offerings [13]. To limit technical complexity in isolating narrative structure, we have based our design on the PNFG language for interactive fiction [5].

3 Narrative Model

Analysis of game narratives assumes of course a suitable representation. As mentioned in the previous section, we make use of a language for constructing interactive fiction (IF) games, *PNFG* [5]. The IF genre has the advantage of consisting of complex narratives, relatively easily extracted from the game architecture. The PNFG format is preferred over more full-featured languages such as Inform [13] as a simplified, but complete IF language with a well-defined semantics and reduced syntax.

IF Properties. An interactive fiction narrative provides a minimal, typically text-based virtual environment. A player avatar is controlled through textual input forming game commands, and the current or resulting game state is reported by textual output. Gameplay is turn-based, and usually involves exploration of an interconnected series of rooms, wherein the player (avatar) may examine, move, and otherwise manipulate game objects. Appropriate actions unlock or control narrative progress, moving the player from an initial state to either a winning or losing conclusion.

Basic game control flow is straightforward. After initialization, the game waits in an idle state for user input; user commands trigger game actions, which can result in either a game win, loss, or (more typically) a return to the idle state following any post-turn processing. The complexity of game state is an important property to analysis. In IF games critical game state consists of simple object properties, such as a room being lit or unlit, as well as the object containment hierarchy—the location of each object, including the player's inventory. Counters and dynamic object allocation can add further complexity, although games are usually finitely bounded, with a small limits on counters and a fixed maximum number of available game objects.

PNFG. The PNFG language provides a minimal model of IF game structure. Game objects and rooms are defined along with boolean state variables, and containment is internally represented by further boolean state (x (not) in y for all objects x and locations y). User commands invoke code which can set and unset object states, move objects between locations, as well as branch conditionally on object state or location. Syntatic sugar is provided for a number of further common IF constructs, including finite counters, scoping of game commands; more language details can be found in [5].

4 Narrative Analysis

Our overall approach to narrative analysis is illustrated in Figure 1. PNFG game specifications are subject to initial, high-level analysis, which is then used in conjunction with the game's interpretable (compiled "NFG") form as part of an optimized search of the game state space. The subsections below describe the basic search behaviour and our optimization.

As an example of useful narrative analysis, and to give our study a specific analysis goal, we investigate the general problem of computing "winnability." This involves identifying the sequence of commands that bring a player from an initial state to a winning game completion. As well as verifying a fundamental game property (the game should be winnable), finding one or all "winning paths" is an instance of the more general problem of efficiently determining state reachability, and our techniques can be easily abstracted from the winnability goal and applied to other search problems.

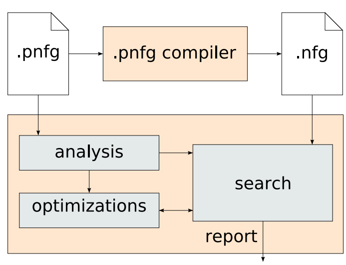


Fig. 1. Overall system design

4.1 Basic Search

The basic state-space search proceeds as a back-tracking, depth-first search of reachable game states, applying all possible game commands at each state. Game state is internally stored in terms of the state of the internal NFG representation, but is not directly accessed by the search system. Reachability and progress are instead determined by the ability of a game interpreter incorporated into the system to apply a given command, monitoring for win, lose or error conditions.

A number of generic optimizations improve performance of this naïve, brute-force approach. Cycle detection is an effective tool for most narratives. Many games, for example, tend to permit a variety of commands such as "look," "examine," or command sequences such as "take x, drop x" that do not overall modify game state, but nevertheless contribute to the branching factor and lengthen search paths. Our system thus maintains a stack of states during its DFS activity, and truncates searches containing states previously encountered in the current game search path. Caching is also used across search branches in order to avoid "dead-ends," or states which have no legal actions that can lead to a game win (i.e., all actions lead to either losing, errors, or dead-ends). These states are cached, and used to further prune the state space search.

Although these optimizations are effective, larger narratives have many commands and objects as well as potentially long solution depths, and for practical analysis a very small branching factor and/or good search heuristics are required. Our approach applies dataflow analysis techniques more typically used in compiler optimization to extract game information in order to further improve performance. Below we describe our primary techniques for optimizing the search process.

4.2 Pre/Post-condition Analysis

To reduce back-tracking in the search it helps to know ahead of time which commands may follow each other. Even if illegal combinations are quickly pruned during search the large branching factor of a naïve search limits scalability.

In the PNFG language, many of the operators that modify game state, either by changing boolean variables or by moving an object from one location to another, assume a specific input state. For example, the command to change a game object variable from false to true (`+object.var`) requires the variable be initially false, and similarly, the object move statement (`move x from y to z`) assumes the object `x` is currently in `y`. Note that successful execution also guarantees the resulting output state: in the first example `object.var` is certainly true if the statement completes without error, and in the second `x` will be in `z`.

These observations drive two basic analyses which we then combine in order to prune the branching factor. We first analyze each action in a *backward* direction, computing a conservative approximation of the minimal necessary conditions for the action to execute correctly. We then perform a very similar analysis in a *forward* direction, conservatively computing the output conditions. In order for one action to follow another then, output of the first must be compatible with the input requirements of the second.

In a formal sense these form symmetric dataflow problems. If we consider our forward, *post-condition* analysis, we associate with each object variable and x, y object location state (x in y) a value from the domain $\{$true,false, $\top\}$; *i.e.,* the (incomparable) elements true and false, along with a greatest element $\top$, representing a state which may be either true or false.

The dataflow technique propagates these pairings through all code paths starting from the action code entry (or exit) with all variables in unknown/inconsistent states ($\top$), modifying the associated domain values according to the game action. The statement `+object.var`, for example, will in a forward sense result in an output pairing `object.var:true`, and in a backward sense the pairing `object.var:false`. At conditionals information is duplicated along each branch, and at join points merged by setting `object.state:`$\top$ whenever `object.var` is not the same on both sides.

A short example of the application of post-condition analysis is shown in Figure 2. In this case post-condition analysis is reasonably effective, determining the location and the state of the candle, although not the state of the player. Pre-condition analysis is symmetric, performed in the reverse direction.

```
(you,take,candle) {                        candle-in-box:⊤, candle-in-you:⊤, candle.lit:⊤, you.hurt:⊤
  move candle from box to you;             candle-in-box:false, candle-in-you:true, candle.lit:⊤, you.hurt:⊤
  if (!candle.lit) {                       candle-in-box:false, candle-in-you:true, candle.lit:false, you.hurt:⊤
     +candle.lit;                          candle-in-box:false, candle-in-you:true, candle.lit:true, you.hurt:⊤
  } else {                                 candle-in-box:false, candle-in-you:true, candle.lit:true, you.hurt:⊤
      +you.hurt;                           candle-in-box:false, candle-in-you:true, candle.lit:true, you.hurt:true
  }}                                       candle-in-box:false, candle-in-you:true, candle.lit:true, you.hurt:⊤
```

Fig. 2. An example of post-condition analysis: dataflow information (on the right) is propagated in a forward direction

Upon completion of this analysis, each action has calculated pre and post-conditions. A variable with a post-condition of true is necessarily true on action exit, and a variable with a pre-condition of true must be true on input if the action is to execute correctly; symmetric for false of course. For each action pair compatibility can thus be easily tested: if the post-conditions of action A include `object.var:`a and pre-conditions of B have `object.var:`b for a given object variable, then action B can follow action A as long as $a \sqsubseteq b$ or $b \sqsubseteq a$. Actions available at each point in the state search are thus selected according to these restrictions.

5 Experimental Analysis

We have implemented our system and examined the effects of our optimized search on a variety of moderate size narratives, roughly the size of a commercial game chapter. In each case we measure the performance of the analysis system as it is applied to a basic winning path problem. Below we describe our benchmark suite, discuss our measurement strategy, and present experimental results and accompanying observations.

5.1 Benchmarks

Benchmarks are drawn from two main sources: the well-known commercial game, *Return to Zork*, where we have modeled the first two chapters in PNFG, and several new narratives developed directly in PNFG by undergraduate and graduate students as part of a course assignment. The latter represent amateur efforts, but were required to respect some basic complexity measures, including lower bounds on number of objects, commands, solution complexity and length. Table 1 summarizes interesting, static properties of our benchmarks. Naïvely, the state graph to be searched will have a branching factor given by the number of commands (global and room-specific), and a depth at least of the solution depth. Assuming any object can be in any room and any room can be in any other room, the overall state space is bounded in size by $|\text{Rooms}|^{|\text{Objs}|} \times (|\text{Rooms}| - 1)^{|\text{Rooms}|} \times 2^{|\text{Vars}|}$.

Table 1. Benchmarks and properties: lines of code, total number of rooms, objects, boolean object variables, commands (global and room-local), and minimum solution depth

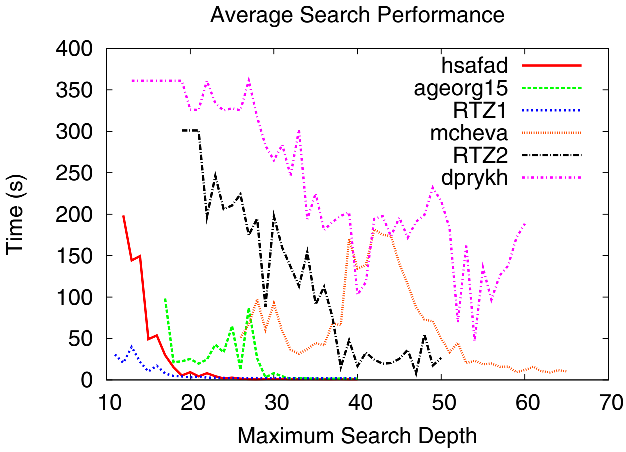
Benchmark	LoC	Rooms	Objs	Vars	Global Cmds	Room Cmds	Depth
ageorg15	1230	9	16	33	15	45	17
dprykh	1472	8	11	23	40	29	13
hsafad	387	10	14	18	2	39	12
mcheva	775	14	10	6	0	62	26
RTZ Chap 1	583	11	20	10	5	42	11
RTZ Chap 2	1113	22	37	16	4	118	19

5.2 Measurements

We measure performance of winnability analysis on our benchmarks using our optimized analysis. Measurement of winnability analysis is complicated by several factors. An exhaustive test, reporting every winning path up to a given depth would, for instance, provide a deterministic workload. Complete enumeration of winning solutions, however, is not practical—the number of possible solutions grows very quickly as search depth increases, resulting in performance being quickly dominated more by reporting time than analysis. We avoid this problem at the cost of greater variance by measuring only the time to find the *first* winning solution. In order to avoid skew introduced by the order in which nodes are examined in the search, we also randomize the order of actions considered as each node is expanded in the search.

Timing results are shown in the graph on the left in Figure 3. For each benchmark and for a range of maximum search depths, a series of 10 analysis attempts are performed and the average time plotted. Note that while general trends are stable variance can be significant, as a fortunate or unfortunate node ordering during search can have a large impact on an individual experiment. All results were gathered on the same quad-core Xeon 2.3GHz machine with 16GB RAM, Debian 2.6.18, using Java HotSpot 1.5.0_14 and a 1.5GB heap, and were limited to 5 minutes per search (except for DPRYKH at 6 minutes). It is important to note that these results represent orders of magnitude improvement over previous work in this area using a generic, SAT-based solver, which was not able to complete given several *days* of execution time for RTZ CHAPTER 1, our smallest (in terms of solution depth) benchmark [5].

Impact of optimization. Our optimized search process has varying performance, but has overall impressive efficiency, finding solutions within our time bounds for all of



Average Branching Factor

	Naïve	Opt
AGEORG15	60	23
DPRYKH	69	44
HSAFAD	41	8
MCHEVA	62	5
RTZ CH. 1	47	10
RTZ CH. 2	122	10

Fig. 3. Search performance on benchmarks

our benchmarks at most search depths. This is particularly evident at larger depths. The DPRYKH benchmark poses the most significant challenge for our analysis, mainly due to the high number of player commands that remain even in the optimized search cases, and hence large branching factor. This can also be seen in the table on the right of Figure 3, which shows the average branching factor for naïve and optimized searches. Optimization is able to reduce this dramatically for most benchmarks, but for DPRYKH the use of many global actions (see Table 1) limits the effectiveness.

Inherent winnability. A striking, and initially counter-intuitive feature of all our data sets is the way search times tend to decrease as maximum search depth is increased. Since the state space is exponential in the search depth, our expectation was that increasing depth would uniformly increase search time. The data in Figure 3 show this is not the case, and we may even conclude that merely increasing search depth is a feasible search optimization strategy, if minimal solutions are not required.

This behaviour can be explained in relation to a few main factors. In a general sense finding a *minimum* length solution is a harder problem than just finding *any* solution. Minimum length solutions tend to be few, and the number of solution permutations grows very quickly if lattitude is given to contain extra sub-optimal actions or orderings. This may increase the relative solution density as depth grows, and patterns in solution densities can cause periodicity in the search performance as well, partially explaining the less monotonic search performance of MCHEVA and AGEORG15. For smaller narratives, such as RTZ CHAPTER 1, our optimized search may also have reduced the state space to the point where it could be feasibly exhausted. Although this is unlikely in general, and certainly not true for larger games, we note that most IF games are both finite and designed to be eventually won. Narrative games are often time-intensive and have relatively low replayability—good game design avoids the need for the player to save and reload earlier states, and thus a player who plays the game without repetition and without losing should always be able to eventually win.

6 Conclusions and Future Work

Narrative analysis has many potential applications. Here we have investigated and shown the practical feasibility of a basic reachability problem, but pre/post-condition analysis is relatively independent of goal and our design could be easily extended to apply to a variety of interesting narrative verification questions. Properties more generically depending on game paths or state relations, such as "pointlessness" [4], or constraints on spatial locality [1] are search-related problems, and could be examined using our techniques. Further application is found in determining the path to more immediate game goals (such as how to open the next door), where search speed improvements are essential for developing online and interactive but generic and automated game hint systems.

Acknowledgements. This research was supported by the Natural Sciences and Engineering Research Council of Canada.

References

1. Nelson, M.J., Mateas, M., Roberts Jr., D.L., Isbell, C.L.: Declarative optimization-based drama management in interactive fiction. IEEE Computer Graphics and Appl. 26(3), 32–41 (2006)
2. Barros, L.M., Musse, S.R.: Towards consistency in interactive storytelling: Tension arcs and dead-ends. Comput. Entertain. 6(3), 1–17 (2008)
3. Adams, E.: The designer's notebook: Bad game designer, no twinkie! parts I–VI (1998-2005), http://www.gamasutra.com
4. Verbrugge, C.: A structure for modern computer narratives. In: Schaeffer, J., Müller, M., Björnsson, Y. (eds.) CG 2002. LNCS, vol. 2883, pp. 308–325. Springer, Heidelberg (2003)
5. Pickett, C.J.F., Verbrugge, C., Martineau, F.: (P)NFG: A language and runtime system for structured computer narratives. In: GameOn'NA: Proceedings of the 1st Annual North American Game-On Conference, EUROSIS, pp. 23–32 (August 2005)
6. Burg, J., Boyle, A., Lang, S.D.: Using constraint logic programming to analyze the chronology in "A rose for Emily". Computers and the Humanities 34(4), 377–392 (2000)
7. Lindley, C.A., Eladhari, M.: Causal normalisation: A methodology for coherent story logic design in computer role-playing games. In: Schaeffer, J., Müller, M., Björnsson, Y. (eds.) CG 2002. LNCS, vol. 2883, pp. 292–307. Springer, Heidelberg (2003)
8. Natkin, S., Vega, L.: A Petri net model for computer games analysis. International Journal of Intelligent Games & Simulation 3(1), 37–44 (2004)
9. Brom, C., Sisler, V., Holan, T.: Story manager in 'Europe 2045' uses Petri nets. In: Cavazza, M., Donikian, S. (eds.) ICVS-VirtStory 2007. LNCS, vol. 4871, pp. 38–50. Springer, Heidelberg (2007)
10. Araújo, M., Roque, L.: Modeling games with Petri nets. In: DiGRA Conference, London, Brunel University (September 2009)
11. Tessman, K.: The Hugo Book—Hugo: An Interactive Fiction Design System. In: The General Coffee Company Film Productions, 1st edn., Toronto, Canada (2004)
12. Wild, C.: ADRIFT: Adventure Development & Runner—Interactive Fiction Toolkit, version 4.0 manual (2003), http://www.adrift.org.uk
13. Nelson, G., Seebach, P., Firth, R., Plotkin, A., Short, E.: Inform. (1993), http://www.inform-fiction.org/

Cultural Computing – How Can Technology Contribute the Spiritual Aspect of Our Communication?*

Naoko Tosa

Kyoto University, Academic Center for Computing and Media Studies.
Yoshida Nihonmatsu-cho, Sakyo-ku, Kyoto, 606-8501, Japan
tosa@media.kyoto-u.ac.jp

Abstract. The author is carrying out technology studies to explore and expand human emotions, sensibility, and consciousness by making innovative use of artistic creativity. We develop interfaces for experiencing and expressing the "essence of culture" such as human feelings, ethnicity, and story. History has shown that human cultures have common and unique forms such as behavior and grammar. We suggest a computer model for that process and a method of interactive expression and experiencing cultural understanding using IT called "cultural computing". We particularly examine Japanese culture, although it is only a small subject of computing.

Keywords: Communication, Software, Information System, Cultural Computing, Interactive Art.

1 Introduction

“The ethnic crisis, the urban crisis, and the education crisis are interrelated. If viewed comprehensively all three can be seen as different facets of a larger crisis, a natural outgrowth of man’s having developed a new dimension – the cultural dimension – most of which is hidden from view. The question is, How long can man afford to consciously ignore his own dimension?” This phrase is from “The Hidden Dimension” by cultural anthropologist Edward T. Hall, 1966.

Nowadays, computers play an important roll in various ways in our life. Cellular phones, e-mails, websites, games as well as PCs are almost parts of our life, and they became the daily items or media. Computers were only ‘machines to calculate something’ at first, but now they are ‘media for thinking and memorization support.’

Let us see the relation between traditional customs and computers. Computers are typically used for calculation to restore something or for historical simulation. Archiving the fading cultures with using computers are barely in the use of thinking and memorization support, but it isn’t an effective application of the ability of computers which now treats multimedia and are connected in network. The present ages often communicate with someone who has another cultural background, so they are needed to understand the history of their culture and other cultures. Because the typical way

* Published in ’10 Spring Forum of Society of Automotive Engineers of Japan, May 00, 2010.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 232–239, 2010.

to understand them is to read the books or to go to the museum, understanding another culture with picking appropriate information is not so easy. Can we understand the culture using computers as 'media for thinking and memorization support'which became more suitable for network, mobile, and two-way communication by the development of information technology? This paper describes basic methodology of cultural computing, that is, to treat the essences of the deep-inside culture like sensitivity, national traits, or narratives. And it integrates them into verbal and nonverbal information, proposing the prosperity of the field that treats the experience of exchange between cultural experiences and culture models by using computer. The cultural computing, which is essential for the communication ability of future computer, introduces you to this new field which defines what humans have stored in each culture and it's history in forms of actions or grammars are sharing common or peculiar forms by showing some concrete methodology and some examples.

2 Media for Thinking and Memorization Support

2.1 The Transcending Artist

What is the aim of arts? This is very huge and heavy question, but if I dare to answer, it is 'the visualization of the status of someone's heart' in a word.

The value of arts is in contrary of the one of technologies. Great art works are having universal values that have not faded yet. On the other hand, old technologies are often selected and surpassed by other new ones. Media arts does transcend them, making new relationship between art and technology, and affecting to other various fields. In other words, arts transcend technologies and technologies transcend arts.

I was conscious of my way as a transcending artist, and determined to obtain global viewpoint instead of local, Japanese culture as the next step. I moved my base of research activity to Massachusetts Institute of Technology, USA. However, I faced the difference between cultures there. I noticed that there were the differences not only in our daily life, but also in the feelings, memories, sign, unconscious communications that are strongly related to their cultures. I noticed that extremely Japanese-characterized expressions as well as global express had also been used in my works, although I had tried to express my artworks generally.

I have tried to assimilate myself into America at first, but it was difficult and I strongly realized that I am Japanese. I paid attention to actions and grammars that are seemed as 'Japanese culture' in America. I tried to expression the difference between American behaviors or grammars and Japanese ones in the arts.

At that time, I have met Sansui ink paintings (Japanese landscape paintings) of Sesshu. Sansui paintings are not landscape painting, but they are imagined scenery. This is in relation to the unconsciousness of occidental Jung's psychology. I discovered the way of unconscious communication to understand other cultures by using the components of unconsciousness in Sansui paintings.

I succeeded to create an interactive artwork, 《ZENetic Computer》 by modeling the structure of Zen and Sansui, which are thought the very Japanese culture. After that, I created the Kanji inspiration 《i.plot》 which gives relations between psychological associations and graphical images of ideograms. Furthermore, I produced 《Hitch Haiku》 system which supports creating the Haiku from several Kanji (Chinese character) input, using the template of 5 – 7 – 5 characters.

Surprisingly, when I exhibited these works in overseas, for example MIT, many Americans understood these works. This shows that these arts, transcending the time as history or culture, and the space as nations, became the media which people in other culture can understand, by picking up the structure of peculiar traditional Japanese culture through the computer.

2.2 Cultural Poiesis

By continuing these studies, I discovered that I could precede the methods of art and technology. That is, interactive works with computer modeling of Japanese culture can be brand new media that enable everyone to understand other cultures by touching them. Thus, obstacles between Japanese traditional culture that was thought as peculiar and other cultures were slowly disappeared.

To express a core of a culture based on the traditional culture model, by using computer. Many people from around the world can create their own Haiku or Zen, and send them to the world by the computer using traditional cultural model through the metaphor. Of course, computers can model not only Japanese cultural metaphor but also the world theater of Shakespeare, and it is possible to create Kabuki composed by the metaphor of the Globe Theatre. I think this can be called as the creation of the culture, the cultural computing.

I noticed that computers have a feature appropriate to create the new culture. Computer processes are divided into algorithms and data. We can seem them as types and contents! Handling the culture by the computer would lead to the creation of the new culture. The culture in that context is a poiesis (that means creation in Greek) of communication between different cultures.

3 From Occidental Unconsciousness to Eastern Sansui Paintings

3.1 We Cannot Take Our Culture Off

Have you had impatience about communication and thought, "Communications usually succeed, but why do they occasionally fail?" It fails because you communicate with someone supposing that they should be able to understand you. Communications are needed, however, for they cannot understand you. If you get it, you will find how to communicate with them calmly.

If we have understood all of us already, we do not need to communicate with each other. Japanese people has a strong pride of homogeneous race, so we apt to have relatively the same feelings, impressions and opinions about someone's actions, phenomena and a course of our society. As they typically call it a 'direct communication from mind to mind (Ishin-denshin in Japanese)', we are prone to think something supposing that they should understand us.

The characteristic of Japanese becomes obvious when we go abroad. That's because we face the condition that we cannot make ourselves understood even if we have thought they should be able to understand. Many other Japanese looks hard to be normal, like herbivorous animal living with carnivorous animals.

In America, however, especially Anglo-Americans often communicate with considering that a person cannot understand them, I think. Actually, I had colleagues from Germany, Greece, Lebanon, Japanese, Chinese, French and Anglo-America. Southern Europeans usually use a nonverbal communication like Japanese, which makes me feel a sense of closeness. However, Anglo-Americans keep a distance when they communicate. Somehow, they do not open their hearts, or they tend to hesitate at showing their feelings. This fact shows that they communicate with considering that a person cannot understand them.

The same situation sometimes occurs when we communicate with a computer as well as a person.

In Boston, audio response systems often answer when we call to enter into contracts with a telephone or a gas service. Computers ask us, “My name is Alice (e.g.), please answer me to register.” “May I ask your name, please?” “May I ask your address?” Those questions continue. And what is worse, the speech recognition sometimes fail. Then the computer says, “I couldn’t recognize, please repeat that again?” Answering it three times will be the limits of our patience. One of what are good about American frontier spirit is this big-hearted and trendy disposition.

“Do people here really patient about computer operators?” I asked my friend and she advised me , “In that case, you should wait while human operator appears.” She says that various things (computers or humans) meddle in phenomena, and extended the time we decide something. This analysis also shows the difference between Japanese who accustomed to communicate tacitly and Americans who accustomed not to communicate tacitly.

While I stayed at Boston, I realized by my experience that we should think communications occur only when people can’t understand each other, considering the difference of their culture. Many times I experienced satisfaction of sympathy with someone whom I thought was not understandable, which is beyond expression. We cannot communicate freshly without this moment. If we communicate not only with discovering our errors, differences or sympathies but also with exchanging and amplifing our knowledge and feelings, the communication will transchend each cultures.

We obtained global comunication by adopting medica technologies in face-to-face communication which had been limited to a small community. E-mails, social networks, blogs enabled us to communicate more easily with people from around the world, beyond barriers of distance and culture. On the other hand, many people feel communications being more and more shallow these days. Rather, these shallow communications brought a recent typical face-to-face conversation, “Did you read my message?” Communications may be turning into extremely superficial communications, with taking off the tastes of cultures.

Ignoring this tendency will cause the decline in our communication ability which we have had as the basic instinct from ancient times. We immediately need the new communication media which can convey one’s depth of feeling crossing the border of cultures. I knew it is realizable during my 2 years stay in Boston.

I wanted to create the communication media with which we can communicate deep feelings transcending the culture. As the result of my stay in Boston, I had this strong idea.

3.2 Technologies Combined with the Spirit

I visited Western China for 10 days, in the later July 2002. The aim of this trip was to discuss with Tibetan doctors and philosophers, and to complete the fieldwork for my research theme that looks for the problems about arts, technologies and hearts.

Tibetans naturally have an idea that the medicine and the philosophy is the same. Doctors are philosophers, and are Buddhist priests at the same time. I was impressed by the nature that doctors see what is wrong in patient's heart at first.

Tibetan philosophy has an faith to give freedom to all the afflicted lives. Their thought is deeply related to consciousness, feelings, the space and lives, centering the bowels of mercy and wisdom. The cosmic view has to do with our essential problems, the wheel of life and the existence. Many cultures in Tibet include Buddhist Tantrism as the appreciation of the idea that life cycles.

We visited Arura Tibetan Medical Center. We discussed with five people including famous Dr. Denchi, the hierach and Buddhist philosopher, and Dr. Tanjinja who is nyingma (specialists for sadhana in Tibetan cabala). When we referred our spirituality of arts and the possibility of fusion between art and technology, they identified with us and said, "It is a possible idea as one of the figure of future religion." This encouraged us so much.

Tibetan Buddhists set the sprit of bodhisattva very important. This means to put off their tenacity to themselves and self-love, to have altruistic love. I noticed that this spirit is deeply related into interactive art.

Interactivity in art is shallow and the value is low if the purpose is up to self-assertive or communication of feeling. What is important is the interaction with having the sprit of bodhisattva and altruistic love. That is, if computer systems succeed to interact with having the sprit of bodhisattva, the interactivity of the system can deeply resonate the high spirit with other people.

I knew in my visit to Tibet that the Buddhism, which was born in India, has stayed in Tibet adjusting to the climate there and that the global consciousness is remaining there. I was impressed with the fact that we Japanese and Tibetan Buddhist are able to understand each other at a deep side of our spirit. I hoped to create something that Westerners can also understand the Buddhism, as a media expression using technologies. I just met the Sesshu, which critically affected my artworks after that.

3.3 Meeting the Sansui

I met the Sesshu at the exhibition "Sesshu – special exhibition at 500th anniversary of his death" (Kyoto National Museum, 2002). I was fascinated with the world Sesshu had created. I did not have a special interest in Japanese culture before. For some reason, the Sansui world of Sesshu in that exhibition seemed a virtual reality which expressed his heart!

In old China, a Sansui picture was once about a landscape we wanted to watch forever, a place we wanted to go to play, a place we wanted to live, and a hometown of our heart in which we wanted to pass away. The Sansui picture is imagined scenery like that. Its bleeding, cracked, feathering lines of ink brush draw the movement of the heart. It makes us feel the color even if it is monochrome.

I had an inspiration to compute the Sansui picture typically by Sesshu and the world of Zen which was expressed in Sansui picture. Zen makes us feel Japanese culture by its absence of absolutes, beauty sense of "Wabi-Sabi", and getting rid of the water from the Chinese garden with taking the Asian culture in. Many elements in Japanese culture are gathered in Sansui, like the Ume-Sansui (Sansui with Japanese aprlicot). I wanted to express the Sansui picture and Zen culture, centering the Japanese Zen.

Cultures consist of 'the God, the Buddhist image, a view of Life, a view of world'. They have been created, changed, opposed and fused with each other, and are irrational and rational. They, which have both irrationality and rationality, have seemed difficult to handle.

There was no research to make use of the hierarchy of Japanese culture to the computer logic in the existent computer technology. This is the reason why there are no arts expressing the deep historical 'culture' in large scale yet. Another reason is that everyone have paid attention to the uniqueness of Japanese culture and seemed the Japanese culture as Japanese superficial expressions.

On the other hand, technologies finds the mechanism of phenomena, and analyzes the elements of them with the structure. And the study finds new relationship between different things and constructs them, by reconstructing them, trying some combination of them and comparing them.

What we can make use of in creating new media arts are extracting the basic structure or thoughts of Japanese culture, modeling them or using them as tools with using the technology. Fresh media works or art works would likely be created by that. This method will bring about a great possibility to the advancement in media arts and interactive arts hereafter.

Thanks to the cooperation of Seigo Matsuoka, who is a researcher of Japanese Culture in Editorial Engineering Laboratory, we took a little advantage in this difficult challenge that we reconstruct the world of Zen which was expressed in Sansui picture on the computer. Though we needed three years, we reached to the unique system named 《ZENetic Computer》 as the result.

We succeeded to construct the extremely futuristic interactive system by projecting a part of allegory or symbol in Sansui pictures, Yamato-e (Japanese traditional paintings), Haikus, Kimonos that reminds of the Japanese culture – the structure of the oriental thought, the structure of Buddhist philosophy and the mechanism of Japanese traditional culture – which rarely have featured by the computers before.

This system uses various symbols and allegories that are included in Buddhism, an oriental thought and the Japanese culture. This is because they include a plenty of implications, and they have extraordinary terms, figures or colors. There are many rules in Sansui pictures and the world of Zen. We discovered that computers can handle them, if we can select and extract them. For example, there are 'San-En' which is an expression of Sansui pictures, and 'Go-Un' (five elements which form a self-existence) which is a function to recognize the human in Buddhism and so on.

The first exhibition of this system was in MIT museum. I wondered whether Westerners understand it or not. As a result, however, it was accepted by many Westerners and won the great popularity. Westerners had felt that Sansui pictures and Zen is extremely oriental and hard to approach, but they gave me impressions that they could understand them through the interaction with this system. I myself had an impression that we could achieved the initial goal to express Japanese culture in media, when I saw a American child interacting joyfully with this system.

After that, it was exhibited in SIGGRAPH, the international conference of CG, and Kodaiji, a Zen temple in Kyoto. Each exhibition won great popularity. This success of the experiment using this system made me certain that the ‘cultural computing’ which computes the culture is reasonable to set to my research goal.

4 Structure of the Culture Becomes a Communication Technology

4.1 An Interaction to Reach the Racial Memory

I was encouraged by the success of 《ZENetic Computer》, and felt that interaction that reaches the deep-inside racial memory was the research I wanted to realize as the next stage. From the dry interaction of computers to the friendly and impressive interaction. How to realize this challenge?

I tried to classify the types, structures and relationship of what supports racial memories in Japanese culture with my co-researcher, Seigo Matsuoka. Below is the detail:

1. Japanese natural climate: Japanese transient weather and nature, thought of transience like ‘Monono-Aware’, beauty senses like ‘Wabi-Sabi’, existential thought that loves present situation.
2. Relationships between Japanese culture and Asian one : Transformation from Chinese Sansui pictures to Japanese ones, Chinese gardens and grove gardens to the Japanese Rock Garden.
3. The syncretization of Shinto and Buddhism: The cultural structure that was reconstructed as a belief system, mixing the native faith and the Buddhism.
4. Characteristics of Japanese language: Waka poem, Haiku poem, Noh thater, and the script of Kabuki. And as applications, Honka-Dori, Uta-Makura, Kakari-Musubi, etc.
5. Japanese Design: Japanese designs are the most popular. Two-dimensional designs are Mon (armorital bearings), Ori (pattern of textiles), colors, paper patterns, lines for example. Three-dimensional dynamic designs are the design of Noh, Kabuki, etc.

On these bases, we can consult on various racial types of Japanese culture and the rear communication.

4.2 Computers Do Not Have a Cultural Information Hierarchy

Scientific technologies have developed Web2.0 like Google, Youtube, Wikipedia and SNS by which we can send our information more easily. Robot technologies are also developing new basic techniques to realize the superior functions that living things have. More specifically, they are global communication technologies including the movement functionality, manipulation system, distributed autonomous system for upgrading the intelligence.

There is, however, no cultural information hierarchy that is needed to live with humans. Adding the local cultural information by cultural computing to here may contribute to create higher-level communication systems.

5 Conclusion

These methods of 'cultural computing' enables us to model and structure the deep-inside essentials of culture like sensitivity, intuition, racial characteristics and narratives that we have not able to quantification. I have set my goal to realize the communication that moves one's racial characteristic expanding the present computer's communication ability to have an ability to reflect the difference in feelings, consciousness and memories, based on the culture. If these systems are realized, social practical and cultural information expression systems through the languages, voices and movies will be realized in various fields.

References

1. Tosa, N.: Cultural Computing. NTT Publishing (2009) (in Japanese)
2. Stafford, B.M.: Visual Analogy: Consciousness as the Art of Connecting. The MIT Press, Cambridge (2001)
3. Minsky, M.: Society of Mind. Simon & Schuster (1988)
4. Hall, E.: The Hidden Dimension. Anchor (1990)
5. Nakatsu, R.: Future of Communications: How can technology Contribute the Spiritual Aspect of our Communications. Ohm-Sha Publishing (2010) (in Japanese)
6. Tosa, N., Matsuoka, S.: ZENetic Computer: Exploring Japanese Culture. The Journal of the International Society for Arts, Sciences and Technology: LEONARDO 39, 205–211 (2006)

System and Context – On a Discernable Source of Emergent Game Play and the Process-Oriented Method

Barbara Grüter, Miriam Oks, and Andreas Lochwitz

Hochschule Bremen, University of Applied Sciences,
Flughafenallee 10, 28199 Bremen, Germany
{barbara.grueter,miriam.oks,andreas.lochwitz}@hs-bremen.de

Abstract. Mobile games are based on the physical movement of players in a game world, combining real world with virtual dimensions. As the real world defies control, the magic circle, the border of the game world, becomes permeable for influences of everyday life. Neither the players nor the designers nor the researchers are able to foresee and fully control the consequences of players' actions in this world. In our paper we introduce a case study. Within this empirical study the difference between the game as a system on the one hand and the context of play on the other hand becomes discernable as a source of emergent game play. We then elaborate on its meaning for the process-oriented method.

Keywords: mobile games, system, context, process-oriented method, emergent game play, quantitative method, qualitative method.

1 Introduction

Mobile games are based on the physical movement of players within a game world that merges real world with virtual dimensions. This kind of games has been spread meanwhile worldwide embodied by numerous prototypes and accompanied by various debates on the nature of these games: location-based, context-aware, ubiquitous, augmented reality and pervasive [1]. Since 2007 mobile games are becoming commercially relevant, see for example GPS Mission, fastfoot challenge, Parallel Kingdom or locative social media like foursquare and Gowalla. Accordingly the issue of evaluating mobile applications has been put forward [2]. However, evaluating mobile applications researchers face unprecedented theoretical, methodological and organizational issues. Mobile games are situated in the everyday world. As the real world is only partly foreseeable and controllable, players, designers, and researchers of mobile games encounter manifold aspects of everyday life they have to deal with playing, designing and evaluating the game. Similar to Juul [3] we assume each single object, action, and player of game play to be defined by the game system and the context of play (p. 514). We see this difference between system and context as the source of emergent game play, which have to be taken into account methodologically when it comes to study play experiences and to evaluate a mobile game. Today researchers deploy traditional

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 240–247, 2010.

HCI methods and seek to adapt them to the novel situation [4]. Ethnographers around Crabtree and the Mixed Reality Lab Nottingham conceive game play as the result of players' actions and consequently look for the particular methods players employ to organize their play action within their own context [5], [6]. Reid et al. propose an emergence-driven research methodology by combining field trials and experiments to discover emergent facets of game play [7]. In this paper we contribute to these studies by zooming in an actual game play process. We want to make the difference between the game system and the context of play empirically discernable as a source of emergent game play and to elaborate on its methodical meaning for the process-oriented method we use for evaluating mobile games.

2 A Case Study

In 2006 and 2007 we several times play-tested the mobile game *On the Streets* we have had developed as a prototype from 2004 to 2007 Meanwhile up to 500 players played the game. In the following we introduce and compare two different, very brief mobile fight events [8], which happened during two small play-tests, one with Chinese players and one with German players. It might be important to note that the goal of our comparison is not to study and explain the cultural difference, but to demonstrate the difference between the game system and the context of play as a source of emergent game play.

2.1 The Mobile Game *On the Streets*

The mobile game *On the Streets* is played in an urban area divided into squares, we call fields. Players are organized in gangs. A gang consists of one to five runners and a boss who is physically sitting at a Personal Computer in an office and virtually located in the home base of the gang. The goal of the game for each gang is to gain power and influence by capturing as much territory as possible that means fields, and particularly the home base of the other gangs. All runners have a map of the game territory displayed on their PDA's, where they can retrieve detailed information about the field they are currently in as well as about the eight surrounding fields and by which they take action like attacking the opponent.

2.2 The Play Test – Settings and Observations

In the late summer 2007 we organized two particular play tests: one with six Chinese players and one with six German players. With regard to our question here, we found that the cultural difference becomes discernable as a difference between the game system, identical for all players, and the context of play, culturally different for the players. Both games have been played with two gangs, each with two runners and one boss, in the urban area around our university. The playground consisted of 16 fields, each of them 30 by 30 meters.

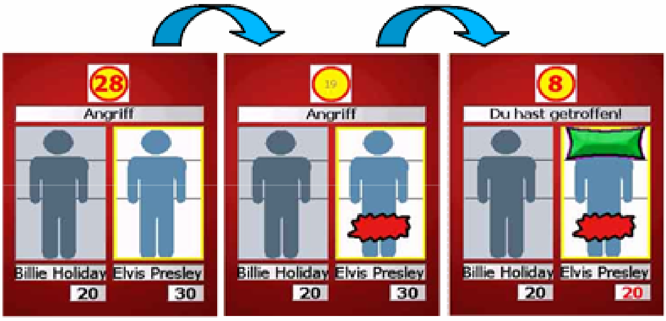


Fig. 1. Fight Mechanics - sequence of fight screens on PDA

The fight mechanics, a component of the game logic, and of the underlying technology, has been used from both player groups. To conquer a field a player has to move into that field. As soon as an opponent player is within the same field each player has the possibility to decide whether he wants to attack or to flee. To attack the opponent they have to identify their enemy physically and to choose the appropriate virtual symbol of their enemy to start the fight.

Fig. 2. The Chinese fight

The fight action enabled by the fight mechanics takes place virtually only. Each fight consists of a sequence of rounds, which a player quits by fleeing. The game logic the players deployed was identical.

However, the players of both groups fought quite differently. The Chinese players fought in a charily manner, kept physical distance to each other and focused on the virtual level. The German players fought in an offensive, expressive manner. They pestered each other, and focused on both, the virtual and the real level, compare Table 1.

Fig. 3. The German fight

We prefer to describe the different behavior and hesitate to explain it. We tended to assign the behavioral difference to the different culture. Meanwhile having discussed the case study with further researchers from different cultural background, German and Chinese, we perceive the different behavior to reflect the culturally different border between the private and the public sphere: while the German players play the game within their native world, within which they do not hesitate to express their emotion directly, the Chinese players play within a foreign world, carefully hiding their emotion.

Table 1. The fights – structure and context dimension of the play activity

	Fight Mechanics - game logic	Fight Mode - context
Chinese	o Enter the field physically o Identify the enemy physically o Attack and defend virtually	o Chary o Kept physical distance o Focus: virtual level
German		o Offensive, expressive o Pestered each other o Focus: virtual & real level

3 The Process-Oriented Method

In the following we briefly introduce the process-oriented method and the way we applied it in our analyses of each play-test, before we particularly elaborate the methodical meaning of the difference between system and context for the process-oriented method.

The goal of the process-oriented method is to study change within the process, within which it happens. We usually do this by means of an iterative cycle consisting of five analytically separable steps, cf. [10]

1. Understand the play process and the steps by which it is unfolding.
2. Identify game events, which either irritate you or change the course of the process.
3. Identify the state before and the state after the irritating or process-changing event.
4. Analyze both dimensions of the play process and the game event included: the goal oriented structure of the play process on the one side, and the varying context of play on the other side.
5. Understand the change in the process by means of the results gained in steps 3 and 4.

In the case study we report here, we applied all five steps iteratively to understand each game play process sequentially.

3.1 Data Collection and Analysis

We aimed for collecting data, allowing us to access the process of the game play: process data streams. We looked for the minimum of these streams, required to understand the game at the level of the whole game, the level of the gang, and the level of the single player: server log-files of all game events, and client log-files of the step-by-step interaction, a re-playable visualization of all runner movements and game logical events during the game play on the map, video-recordings of four runners' play activities, video-recordings of the facial expressions of one boss, Automatic screen-capture of this boss' PC, audio-recordings of the internal gang communications.

Step 1: First, understand the game process as a whole at the level of the game and the level of the gang, watch the recordings several times, particularly the visualization of runners' movements throughout the game. Leadings questions are: who is the winner, how does the game unfold, what are the main phases; how is the game unfolding from the perspective of the different gangs; how does the win/loss ratio change for each gang over time. The outcome of this step is a sequence of meaningful main game phases, often visualized by a time line of events. Second, understand the game process as a whole at the level of a single player: watch the available observation videos of single players, and understand the player's communication with other players. The outcome is a temporal sequence of meaningful phases of the process for the single player.

Step 2: Identify single game events, which are irritating, surprising or differing from the expected or planned behavior from the perspective of the player or of the researcher. In the reported play-tests surprising were the fight events and the different behavior of the German and the Chinese players.

Step 3: Analyze the process state before and the state after of the event. If the single event changes the course of the game you have to understand, what, why and how the change happens. What was the course of the game process and what was the player's intention before the change. What is the effect of the event on the further course? For example in the German play-test reported here: after the observed fight both German players concentrated on fighting against each other again and again, ignoring other play possibilities and own strategies outlined before the event. Another outcome at the research process level: After this case study we changed our idea of social interaction in mobile games. We assumed from the beginning a strong physical dimension of social interaction, what we assigned to empirical phenomena resembling the German fight mode [10]. The Chinese fight mode kind of irritated and surprised us. Their social interaction is characterized by a different physical quality like distance, caution, focus on the virtual level, we have not imagined designing the fight mechanics.

Step 4: Analyze both dimensions of the process, the event included: the goal oriented structure and the context of play. A possible outcome of the structural analysis is a temporal sequence of play actions, which follow the game logic, and in line with player intentions, which of course may also deviate from the game logic. The purpose of the contextual analysis is to understand the varying situations. An outcome is both, analytical categories, which evolve throughout the analysis, and empirical indices of behavioral differences, to describe the unexpected behavior, unexpected by the player or the researcher. As the main purpose of this paragraph on the process-oriented method is to understand the methodical meaning of this fourth element of the five steps we refer to the next section.

Step 5: Describe and try to develop an explanation of the observed behavior by using the results of step 3 and 4, see for example chapter 2 in this paper.

3.2 On the Methodical Meaning of System and Context

In the current paragraph we particularly emphasize one element of the process-oriented method, the step 4, which we applied to compare two fights from different play-tests. The goal of the comparison was to understand how the interplay of the game system and the context of play works. The goal now is to elaborate on its methodical meaning.

The process-oriented method we use to study play experiences and to evaluate mobile games takes into account that time matters: every time a game is played, it is played differently. As Salen and Zimmerman [11] state: "How does pleasure emerge and evolve over time in a game? All of the possible states and experiences of a game are contained within the theoretical construct called the space of possibility. A game player begins his or her journey through the space of possibility at the same place every time: the start of the game. But the experiential path that a player takes through the space will vary each time the game is played. Every play of the game will be unique, even though the rules of the game, its formal structure, remain fixed. This quality of games, that a game provides the *same* consistent structure each time but a *different* experience and outcome every time it is played, is a powerful engine that sustains and encourages play. We refer to this concept by the shorthand term *same−but−different*." P. 340.

What is valid with regard to the *sequential* repetition of a game by one person, in that the person follows the logic of the same processing game system, however, in different ways, is valid in our case with regard to the *parallel* repetition of a game by different persons. They follow the logic of the same processing game system, however, in different ways.

The requirement of the fourth step of the process-oriented method in the case we presented here was to analyze the goal-oriented structure of the process here the fight mechanics sequentially applied by the opponents, and the changing context of play, here the very different fight modes.

The methodical meaning is first that we are able to analyze the game process by means of quantitative methods as far as it is organized by the game logic, second that we have to apply qualitative methods if we want to understand the context of play and third, that the concrete fight event allows to understand how both methods work together.

4 On the Relevance of Studying Emergent Game Play

Players have the power to change the logic of a game. Hardcore players often try to outsmart the system, see for example [12]. And even being a naive casual player you apply the game system to your particular context. Emergent game play is the first step of transforming a game logic. The identical fight mechanics in our case study worked with different cause-effect relations established within the emerging context of play. Repeated over time players develop implicit rituals and rules by means of which they adapt the game logic to their particular context of play. Families playing a card game do that all time. These implicit rules become explicit as soon as an outsider suddenly interrupts the flow and a conflict on how to interpret the rules has to be resolved by negotiation. Operative rules may result in a redefined game within which the old logic becomes a restructured part of a new system implemented either socially and/or technically.

5 Summary and Conclusion

We presented a comparison of two fights, which worked with the same fight mechanics and different play modes and made the difference between system and context of play discernable. We introduced the process-oriented method and elaborated on the methodical meaning of the difference between system and the context of play. We finally pointed to the relevance of studying game play process-oriented.

To study game play processes by means of the process-oriented method allows researchers to understand concretely the power of players inventing play possibilities and even changing a game. To play-test game prototypes and studying the process of play step by step allows the game designer to learn from players. Longitudinal studies of continuously running mobile game play processes will become an important instrument of future game design.

References

1. Montola, M., Stenros, J., Waern, A.: Pervasive Games. Theory and Design. Morgan Kaufmann, San Francisco (2009)
2. Barnard, L., Yi, J.S., Jacko, J.A., Sears, A.: Capturing the effects of context on human performance in mobile computing systems. Personal and Ubiquitous Computing 11(2), 81–96 (2007)
3. Juul, J.: A Certain Level of Abstraction. In: Baba, A. (ed.) Situated Play: DiGRA 2007 Conference Proceedings, pp. 510–515. DiGRA, Tokyo (2007), `http://www.jesperjuul.net/text/acertainlevel/`
4. de sá, M., Carrico, L., Duarte, L., Reis, T.: A framework for mobile evaluation. In: Proc. ACM CHI 2008, April 5-10, pp. 2673–2678 (2008)
5. Crabtree, M., Benford, S., Greenhalgh, C., Tennent, P., Chalmers, M., Brown, B.: Supporting Ethnographic Studies of Ubiquitous Computing in the Wild. In: Proceedings of the ACM Conference on Designing Interactive Systems, DIS 2006, pp. 60–69 (2006)
6. Sherwood, S., Reeves, S., Maitland, J., Morrison, A., Chalmers, M.: Adapting Evaluation to Study Behaviour in Context. International Journal of Mobile Human Computer Interaction 1(2), 37–55 (2009)
7. Reid, J., Hull, R., Clayton, B., Melamed, T., Stenton, P.: A research methodology for evaluating location aware experiences. International Journal Personal and Ubiquitous Computing Theme Issue: Player Experiences in Location Aware Games – Methodological Issues (to be published in 2010)
8. Binder, D., Wang, L.: Chinese and German players playing on the Streets. Collaborative Master theses; Binder, D.: Cultural diversity and its impact on media design. Hochschule Bremen (2008); Wang. L.: Mobile Games: Interfaces and the cultural context of their use. Universität Bremen (2008)
9. Grüter, B.M., Breuer, H., Wollenberg, A.: Genese von Wissen in aufgabenorientierten Gruppen. Eine Fallstudie zur Wissensarbeit in der kommerziellen Softwareentwicklung. In: Witte, E.H. (Hrsg.) Leistungsverbesserungen in aufgabenorientierten Kleingruppen, pp. S149–S179. Pabst Science Publishers, Lengerich (2000)
10. Grüter, B., Oks, M.: On the Physical Dimension of Social Interaction in Mobile Games. In: Magerkurth, C., Åkesson, K.-P., Bernhaupt, R., Björk, S., Lindt, I., Ljungstrand, P., Waern, A. (eds.) PerGames 2007. Proceedings of the 4th International Symposium on Pervasive Gaming Applications, Salzburg, Austria, June 11-12 (2007), `http://www.informatik.hs-bremen.de/gob/daten/texte/grueter_oks.pdf`
11. Salen, K., Zimmerman, E.: Rules of Play. In: Game Design Fundamentals, The MIT Press, Cambridge (2004)
12. Aarseth, E.: I Fought the Law: Transgressive Play and the Implied Player. In: Baba, A. (ed.) Situated Play. Proc. DiGRA 2007, September 24-28, pp. 130–133. The University of Tokyo, Tokyo (2007)

Re-envisioning the Museum Experience: Combining New Technology with Social-Networking

Madhuri Koushik, Eun Jung Lee, Laura Pieroni,
Emily Sun, and Chun-Wei Yeh

Carnegie Mellon University
Entertainment Technology Center
Silicon Valley Campus, Spring 2010
madhuri.koushik@gmail.com, ejleeart@gmail.com,
lpieroni@andrew.cmu.edu, emilysun@andrew.cmu.edu,
glaciedge@gmail.com

Client: California Academy of Sciences

Abstract. The goal of the project was to design an integrated system for the California Academy of Sciences that combined new technology with a social-networking based website to promote educational learning. Five mini-games were developed for the iPad and connected to a series of websites through a database. The use of new technology drew in users that would not have otherwise engaged in the experience. Connecting with a social-networking website opens up many future possible implications for expanding edutainment.

Keywords: edutainment, new genres of entertainment technology, social-networking, enhanced museum experience.

1 Introduction

Museums have been expanding to incorporate technologies that allow visitors a more interactive experience [1]. Interactivity not only further engages users, but also provides the opportunity for continued learning. Museums like the Exploratorium enhanced the museum experience using RFID tags [2]. The San Jose Tech Museum uses barcodes on their admissions tickets so that visitors can view photographs and the activities they participated in on a website at home. More recently, the Cooper-Hewitt, National Design Museum distributed to each museum guest an iPod Touch that provided access to interviews and slideshows relevant to the exhibit. While these museums have used these technologies to further enhance the museum experience, there has not been adequate research into what effect the technologies themselves have in drawing in guests.

At the same time that museums have been expanding their connection to virtual information, a recent surge in social-networking games has connected people through websites. In November of 2009, Facebook-game creation company, Zynga, reached 100 million unique visitors per month. Social-networking games have been especially

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 248–253, 2010.

successful because of the inherent sociability, spontaneity, narrativity, and playfulness framework used [3]. While there has been research into the possible use of Facebook for higher educational purposes [4], there has not been focus on creating educational social-networking sites.

Thus far, museums have not capitalized on the success of social-networking sites, nor tested the boundaries of using consumer-hyped technology. The goal of the project was to design an integrated system for the California Academy of Sciences that combined new technology with a social-networking based website to promote educational learning.

2 Overview of System

Although the demographic of guests visiting the California Academy of Sciences varied largely, we decided to focus our target demographic to 6th and 7th-grade students.

The experience begins when museum visitors create profiles on the California Academy of Sciences website. Initially they are able to personalize a limited number of characteristics of their avatars. Once they visit the museum, they play mini-games on iPad kiosks to accumulate points on their accounts. These points can then be redeemed at home by returning to the California Academy of Sciences website. Accessing the website from home allows the user to further personalize an avatar, learn more facts, and compare their scores on the mini-games and their avatar with those of their peers. Points can be redeemed to upgrade the avatar's available attributes and uniforms. The general system architecture can be seen in Fig. 1.

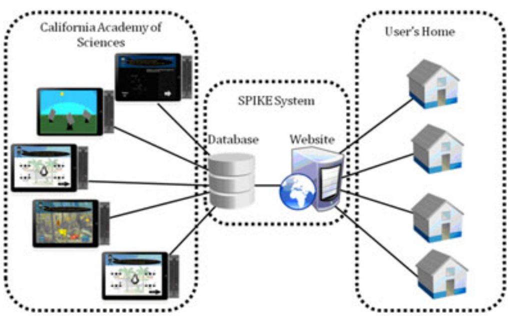


Fig. 1. System architecture connecting data collected from the mini-games in the Academy to the website accessible from home

2.1 Mini-Games

We created five mini-games and developed them for the iPad. The release date of the iPad was April 3, 2010, and the opening of our project was scheduled for April 22, 2010. Based on the popularity and demand for previous Apple products, we anticipated that releasing our games on the iPad, a new technology, would increase the interest in our product. We also wanted to take advantage of the iPad functions, like the accelerometer, that would increase the interactivity of our games.

The five games were placed in relevant areas of the California Academy of Sciences and contained educational material inherent in the games. For the game in the Aquarium, museum guests were presented with a variety of photographs of camouflaged animals and had to tap the animal before time ran out. The name and a fact about each animal were displayed.

In the African Hall mini-game, guests get the opportunity to design their own penguin by altering its height, weight, fin length, and color and see how it fared in the South African wild. Penguins could either be eaten by a shark, drown, or survive. The purpose of this game was to educate students about the difficulties that penguins face every day.

The Altered State Exhibit had guests tap the iPad to move the character to collect the garbage. Tilting the iPad shifted the garbage to allow for easier collection. More points were awarded for collecting toxic waste.

After watching the Planetarium show, guests were invited to test their knowledge of the constellations by finding them in photographs of the night sky. Each game would give the guest the opportunity to find one of six constellations after briefly being shown the constellation (see Fig. 2).

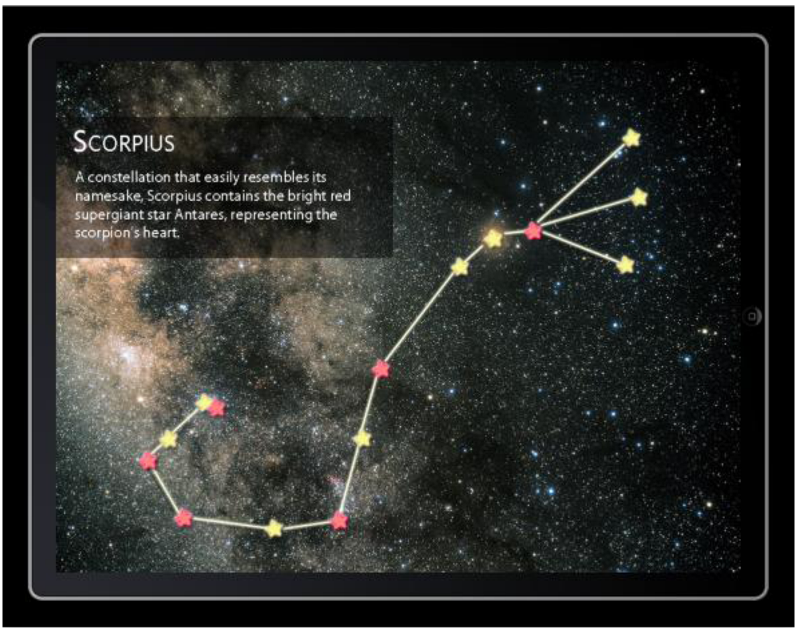


Fig. 2. Example of constellation game and educational fact

Guests guided the blue morpho butterfly in the Rainforest game. By tilting the iPad, users controlled the butterfly's height and had to avoid the arowana fish as well as the silver-beaked tanager bird. They gained extra points by collecting fruit along the way.

At the end of each of the games, the total points and a "did you know" fact were displayed.

2.2 Social-Networking Website

Fundamental to many social-networking games is the creation of the personal identity that represents the user in the virtual world. For the avatars for the social-networking aspect of our project, we had users create their own personal explorer that could have different science-related occupations such as diver and botanist (see Fig. 3). Users redeemed their points from the iPad games to unlock outfits and other attributes for their explorers.

To enforce the connection between the website and the mini-games, a different explorer appeared in each of the mini-games in the California Academy of Sciences. Also, the discover tab was divided into the different areas of the museum where the iPads were located. Within each tab, users were able to find more information relating to the exhibit.

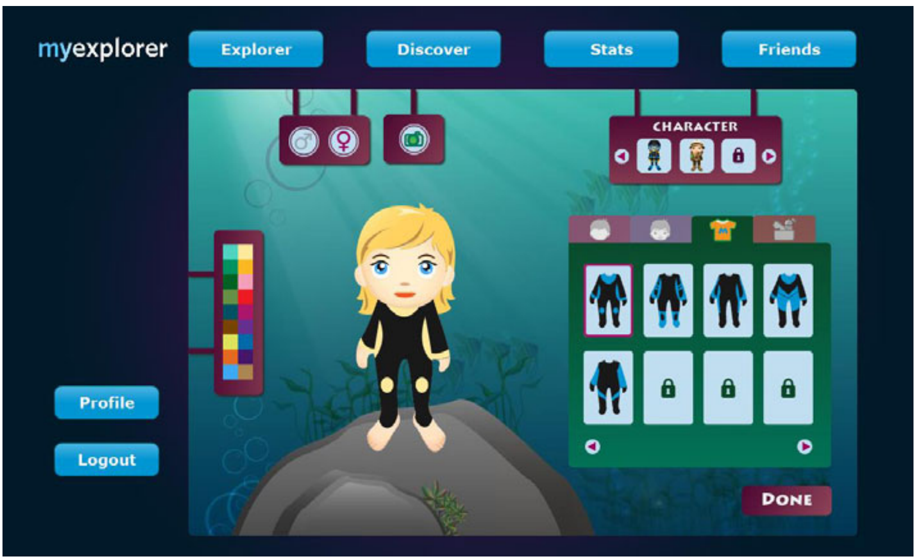


Fig. 3. Screenshot of avatar creation portion of social-networking website

On the website, users were able to add their friends and compare their explorers with those of their peers. The "Stats" tab showed the names and explorers of the users with the top three highest scores on each of the iPad games, as well as the user's highest score and rank on each of the games.

3 User Response

User testing was conducted at a school on 50 students ranging from fourth to eight grade. Students were brought in groups of five and each played a different game on the iPad (see Fig. 4). 72% of the students said they would be interested in redeeming their points online. They also had the opportunity to write a fact that they learned from the game. 67% of students that played the camouflage game (n=12) were able to state a fact that they learned from the game. Example responses to this prompt for all the games include "I learned to make your penguin black or it will be eaten by a shark," "the bird is the enemy of the butterfly," and "animals, such as lizards, camouflage themselves."

Fig. 4. Photographs from playtesting session at school

From our opening on April 22, 2010, we had over 300 museum guests play with the iPad games (see Fig. 5). At both the user testing session as well as opening, much of the excitement came from the fact that they were getting to use a new technology. Many museum visitors would play the game based on the sole desire to try out the iPad.

Fig. 5. Photograph from the opening at the California Academy of Sciences

While it is unreasonable to expect museums to pour their resources into always maintaining the most current technological inventions, capitalizing on the opportunity to develop on commercially hyped technology could jumpstart interest. For something like a socially-networked website, the most difficult part is getting users to initially engage and create a profile. Users are more inclined to join a website that already has a wide following and even more so if their friends are already playing on the website. Once users have spent time and effort on the site, they are more likely to return. Utilizing new technologies is an opportunity to increase the number of initial users that create profiles on a new social-networking game website.

4 Conclusion

This project combined the use of mini-games on iPads in the California Academy of Sciences with a social-networking website to create an engaging, educational experience. There is much more research that can be done to determine the extent of the educational effect of a system like this. For example, how might the experience of learning at the museum, enforcing that learning at home, and repeating through return visits to the website affect retention of facts? As interactive devices and social-networking gaming continue to expand, museums should take advantage of the technology as an opportunity to further educate visitors. It is our hope that this research spurs discussion on the value of educational research in museums and the.

Acknowledgments. This work was sponsored by the California Academy of Sciences.

References

1. Martin, J., Trummer, C.: Personalized multimedia information system for museums and exhibitions. In: Maybury, M., Stock, O., Wahlster, W. (eds.) INTETAIN 2005. LNCS (LNAI), vol. 3814, pp. 332–335. Springer, Heidelberg (2005)
2. His, S., Fait, H.: RFID enhances visitors' museum experience at the Exploratorium. Communications of the ACM (48), 60–65 (2005)
3. Jarvinen, A.: Game design for social networks: interaction design for playful dispositions. In: ACM SIGGRAPH Symposium on Video Games, pp. 95–102. ACM, New York (2009)
4. Mason, R., Rennie, F.: E-learning and social networking handbook: resources for higher education. Routledge, New York (2008)

Interactive Environments: A Multi-disciplinary Approach towards Developing Real-Time Performative Spaces

Nimish Biloria

Hyperbody, Faculty of Architecture, TU Delft, Julianalaan 134
2628 BL Delft, The Netherlands
N.M.Biloria@tudelft.nl

Abstract. The research paper exemplifies upon a series of real-time information exchange driven design-research experiments conducted by the Hyperbody research group (HRG), Faculty of Architecture, TU Delft. These interactive spatial prototypes, while successfully integrating the digital with the physical domains, foster multiple usability of space and are appropriately termed as 'The Muscle Projects' based on the pneumatic muscle driven actuation technologies used per project. The interactive nature of the projects is realized through harnessing a synergistic merger between the fields of ambient sensing, control systems, architectural design, pneumatic systems and computation (real-time game design techniques). The prototypes are thus visualized as complex adaptive systems, continually engaged in activities of data-exchange and optimal augmentation of their morphologies in accordance with contextual variations.

Keywords: Interaction design, Multi-disciplinary, Real-time data exchange, Control systems, Dynamic structures.

1 Underpinning

The design-research experiments developed at Hyperbody, TU Delft, Faculty of Architecture focus on the emerging inter-disciplinary field of Interactive Architecture focusing on developing real-time information exchanging architectural spaces. These interactive spaces demonstrate a fusion between the material, the electronic and the digital domains. This fusion is explicitly attained through harnessing a synergistic merger between the fields of ambient sensing, control systems, ubiquitous computing, architectural design, pneumatic systems and computation (real-time game design techniques). The resultant spaces, visualized as complex adaptive systems, are continually engaged in activities of data-exchange and interact via physical and ambient adaptations of their morphologies in response to sensed contextual variations.

A strategic co-evolution of technical knowledge between the industry (specifically Festo, a pneumatic engineering company), praxis (ONL a multi-disciplinary design firm, The Netherlands), and academic research (Hyperbody for spatial and information design) gives shape to these interactive constructs, hence developing an information bridge between these three critical knowledge sectors. The collaboration process

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 254–261, 2010.

involved a series of associative brain storming, simulation and prototype testing sessions focusing upon the usage of pneumatic and electro pneumatic technologies, interaction design concepts, deciding upon appropriate control systems, structural stability and performance aspects concerning the conceptualized spatial configuration of the system. Experimentations with material systems as regards their flexibility, shape retention and strength ratios coupled with kinetic structural systems formulated a vital part of the collaborative effort. Parallel research and development in interaction design, ubiquitous communication, creating computational routines via software such as Virtools and Max MSP as well as developing project specific sensing networks were also carried out under these collaborative design initiatives.

A systems thinking approach with an emphasis on 'Interdependence and Interconnection' between variable contextual parameters (global) as well as the components formulating a prototype (local) is embarked upon as a part of the synergistic merger of the aforementioned fields. This implied that even the smallest detail in any prototype, for instance every node/joint is conceived as a potential information hub and has the ability to collect, process and communicate contextual data. Apart from this, the same node/joint, via mechanical actuators, is able to kinetically re-position itself physically in three-dimensional space, thus enabling local structural adaptations, which cumulatively result in complex morphological variation.

2 System Components

Issues such as multiple usability of space via spatial augmentation (at a functional level), ambient augmentation (for suggesting the mood/emotional state of the space) as well as information augmentation (as regards the occupancy as well as attraction of the space) attained paramount importance for creating the real-time interactive prototypes. A detailed list of the system components, further segregating them in accordance with their generic usage led to the identification of the following:

a. Pneumatic Entities: Fluidic Muscle Type MAS (provided by Festo): A flexible tube with reinforcing fibers in the form of a lattice structure for up to 10x higher initial force compared to a cylinder of identical diameter. The muscles (Fig 1) tend to contract 20 percent of their initial length with the induction of air pressure, hence making it act as an actuating device to alter the node positions of the prototype.
b. The Black Box (by ONL and HRG with Festo air valves and switching components): A hard-edged box housing the switching mechanisms: I/O boards connected to the 72 valves controlling the air pressure lock of the Fluidic muscles. The box has provisions to attach the compressed air intake pipes through distribution channels; houses the CPU and power back up mechanisms (Fig 1). At a later stage the Black Box was replaced by Festo's CPX controllers (Fig 1).
c. Flexible Skins (utilized in different prototypes): Hylite panel, a sandwich sheet comprising two thin aluminum layers with a plastic core in between (Fig 1) which integrates high flexural stiffness and extreme lightness. Lycra based stretchable fabric (Fig 1) normally used for sports clothing whose translucency varies according to the degree of stretching is also used within some prototypes.
d. Control System: Proximity sensors for sensing the distance of the occupant from the installation, Touch sensors for sensing the amount of pressure exerted upon a surface,

Tilt sensors for determining the x,y,z axis displacements. MIDI, Digital to Analogue and vice versa converters: connecting the sensor input channels and the actuator output channels to and from the CPU. PCI cards, are programmed to receive the output signals (a long numeric string corresponding with the number of fluidic muscles, indicating the status shift of each node: on/off) and communicate them to the In/Out board mentioned earlier (black box), hence controlling the airlock valves.

e. Real-time Game-Design Software: Virtools Dev 3.0, software is used for developing an inherent connectivity between the sensed data and the expected behavior output from the prototype (by means of programming output rules for the system). The software is used as the main computation tool which receives inputs from the MIDI device (sensed data), processes data in accordance with the scripted behaviors programmed into it and sends output digital signals via PCI cards device, which are directly linked with actuating mechanisms.

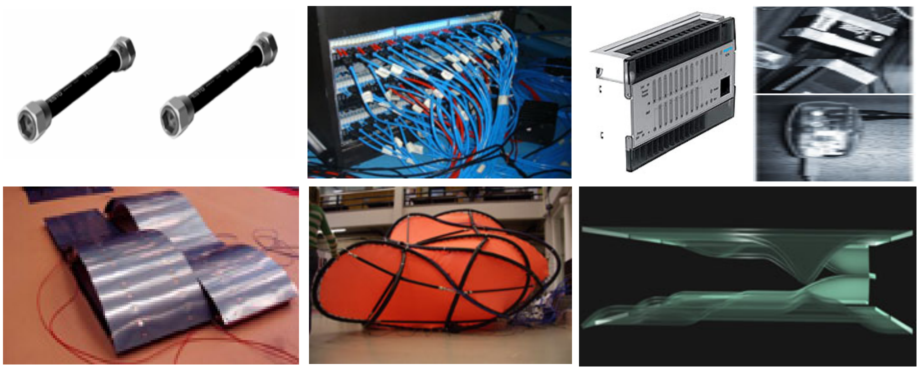


Fig. 1. Components used in the prototypes: Top left to right, Fluidic Muscles, Black Box, CPX controller, I Cube MIDI digitizer, and the bottom row showing left to right the Hylite panels, Lycra skins and the Neoprine envelope.

3 Prototypes

At Hyperbody, a series of prototypes to study the intricacies of research driven design of interactive architectures have been developed over the past six years. The first in the series called the NSA Muscle was specifically built for the NSA (Non Standard Architecture) exhibition in Centre Pompidou, Paris, in 2003 as a collaborative project between ONL, Hyperbody and Festo.

3.1 The NSA Muscle

The NSA Muscle is a pro-active inflated space, its surface populated with a mesh of 72 pneumatic muscles, which were all addressed individually by means of regulating the amount of air pressure induced within them. The prototype is programmed to respond to human visitors through its sensing, processing and actuating enhancements.

Design principles: To communicate with the observers, the NSA Muscle has to transduce physical quantities into digital signals (sensors) and vice versa (actuators). People connect to the NSA Muscle by 24 sensors attached to reference points on the structure. These input devices convert the behavior of the human players into data that acts as the parameters for changes in the physical shape of the active structure and the ambient soundscape. The input setup comprises eight sensor nodes (Fig 2) with three sensors each: motion (for sensing the presence of possible players from a distance of 6 meters), proximity (for sensing the distance of the players to the NSA Muscle within a distance of 2 meters) and touch (for sensing the amount of pressure applied upon the surface). The analogue sensor input channels are converted to digital audio signals (MIDI) and transferred to the computer.

The NSA Muscle is programmed to behave within predefined bandwidths of emotional modes including jumping (excited), retracting (scared) and shivering (anger) which are attained via tactile variations of its external form (by changing the length of the tensile muscles) accompanied by the emission of pre-defined sounds of variable pitch. A three-dimensional visualization (Fig 2) of the Muscle rendered on a computer screen informs the people about the state of this being. The activity of the muscles is displayed in three colors in the model: red / inflating state, blue / deflating state, and gray / passive state, and in the internally used organizational 72 digit string. Images of practical architectural applications, using muscle technology, complement the graphical display. The real time model is actively viewed from multiple camera positions so as to feel the behavioral patterns at work. Viewed in combination with the physical model this graphical interface contributes to the public's level of understanding.

User evaluation: The Muscle NSA owing to its emotional modes, proved to be an attraction for many users. The jumping mode, apart from being activated when many people were interacting with it, was also programmed to activate when the sensing field could not track people around it. This was treated as a strategy to seek attention of the passers by and invite them to interact with the Muscle. Once attracted, the users could also sense the shivering mode (accompanied by a low pitch sound) followed by the retraction of the nodes of the Muscle (at a local level) with which people were interacting. This behavior engaged people in a game play between architectural space and their own emotional selves, thus aiding in changing the perception of architectural space from a static object to a dynamic subject.

Fig. 2. Left: Visitors interacting with the NSA Muscle sensor nodes at the Centre Pompidou, Paris, 2003, Right: The NSA Muscle's real-time computational process being shown via a monitor at the NSA exhibition. The activity of the physical muscles is displayed in the digital world in real-time in three colors in the model: red / inflating state, blue / deflating state, and gray / passive state, and in the internally used organizational 72 digit string.

3.2 The Muscle Re-Configured

The Muscle Re-Configured project was specifically aimed at materializing a real time responsive variant of the Muscle NSA project by ONL and HRG. The reconfiguration is realized by means of utilizing the same actuating components: pneumatic Festo fluidic muscles from the Muscle NSA project. However for the new installation, instead of the soft volumetric alterations of the external form (as was materialized through the Muscle NSA project), an approach emphasizing internal spatial response is visualized.

Design principles: In order to experiment with an interior space, the re-configured prototype was conceived as a 3D habitable Strip (Fig 3): a three dimensional section in space, programmed to respond to its occupants through its sensing (proximity and touch sensors), processing (graphical scripting for real-time output) and actuating (fluidic muscles) enhancements. The construct utilizes the properties of the Hylite panel's being bendable and the fluidic muscle's being able to produce compression power to interact upon each other, thus transforming the otherwise hard edged (visually) spatial strip into soft luxuriant variations. Each Hylite panel coupled with two fluidic muscles hence forms the basic unit of the strip. Subsequent panels are joined together to create a closed 3D loop, in the process creating a series of nodes (junctions where the panels join). These nodes, owing to the possession of actuation members, are linked in space in a highly interdependent manner, constantly exchanging information (in terms of air pressure variations), yet, behaving as a collective whole to attain spatial reconfigurations. The collective whole as such constitutes the following: the responsive floor, ceiling and walls:

The floor units (Fig 3) are strengthened by means of adding wedge shaped Styrofoam sections underneath the inner surface of Hylite panels for seating and relaxation purposes. They also embed touch/pressure sensors (manual control) which, when pressed produce curvature changes by means of varying the amount of air pressure released to the fluidic muscles of the seats. The floor units also embody proximity sensors at their base, which, instantly after sensing the proximity of people trigger the seating units to a pre-programmed position, to create suggestive usability scenarios. The positioning/orientation of the panels constituting the ceiling unit's (Fig 3) is directly related with their affordance (functional and psychological). The ceiling units behavior involves the creation of projection surfaces, generation of smooth curvilinear soothing forms for relaxation purposes and for materializing openings in ceiling surface for allowing light to venture through. These behaviors are attained by means of utilizing the compression power of the corresponding sets of fluidic muscles attached to the Hylite sheet units. The ceiling's tactile variations specifically deal with the users input, specifically the touch sensor readings obtained from the seating units; The ceiling units directly above the seating units are bound together in a way that any curvature variation in the seating units shall have a corresponding curvature variation in the ceiling units. Pressure exerted on the touch sensors in the seating units, therefore creates a subsequent change in configuration of the ceiling units. The wall units (Fig 3) constitute the same generic Hylite panels, which are woven together to create a continuous surface with the ceiling elements. The same principle of compression strengths goes into materializing the wall, which, when actuated bends to create projection surfaces and higher seating surfaces (Figure 3). The actuation of the wall and ceiling elements are also intrinsically linked up with the seating element actuations, hence weaving the entire construct into a cumulative whole. However there are also provisions in which, for experimental reasons, one can individually trigger the three entities.

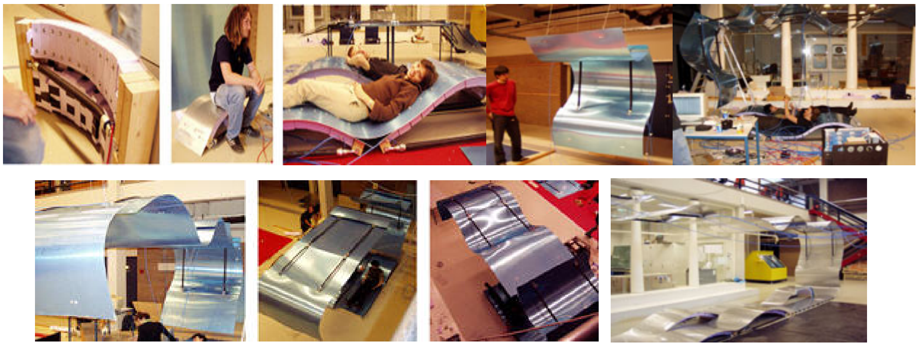


Fig. 3. The three spatial elements constituting the 3d loop of the Muscle Reconfigured; Floor unit's construction and usability scenarios, Ceiling units (displaying controlled opening behavior), Wall units bending to create projection surfaces.

3.3 Emotive InteractiveWall

The Emotive InteractiveWall (Fig 4) is composed of 7 separate wall pieces (herein referred to as nodes) that display real time behavior by swinging back and forth, displaying patterns of light on its skin, and projecting localized sound.

Design principles: The primary synchronous behavior of the InteractiveWall is movement. The nodes of the InteractiveWall will bend independently of neighboring nodes in response to the presence of a user. Although responsively independent, each InteractiveWall node synchronizes by constantly readjusting its position in order to align itself with the position of its nearest neighbors.

Augmented modality of the InteractiveWall's behavior is light. The skin of each InteractiveWall is covered by a unique, irregular distribution of dynamically controlled LEDs that form a highly reactive interface. The LED skins respond directly to user presence by glowing brighter when users are near, and dimmer as they move away. In addition to dimming, the LED skins pulse rapidly and slowly in relation to node position, having a tendency to flash together when the nodes are in sync.

The third modality of the InteractiveWall is localized sound. Moments of synchronicity are represented by calmer sounds, while asynchronous behavior results in more intense sound. The propagation of the sound from high to low intensity is varied throughout the InteractiveWall, thus each node is a member of a choir that sings a complex pattern of oscillating chords. Although similar, the physical movements of InteractiveWall, and the light and sound patterns change independently, reacting at varying rates. The synchronous behavior between the InteractiveWall nodes contrasts with the behavior produced by user presence, resulting in a series of complex wave patterns that propagate through the InteractiveWall structure as a whole.

User experience: The users interacting with the Emotive InteractiveWall experience the installation as an interactive system that exhibited emergent behavior and performed like a living system. The wall worked as an independent system built on synchronous behavior that is interrupted by the game-like response of multi-participant interaction. This layered system encourages the intended cycle of observation, exploration, modification, and reciprocal change in the participant, reinforcing believability in the system, and providing a sense of agency to the user.

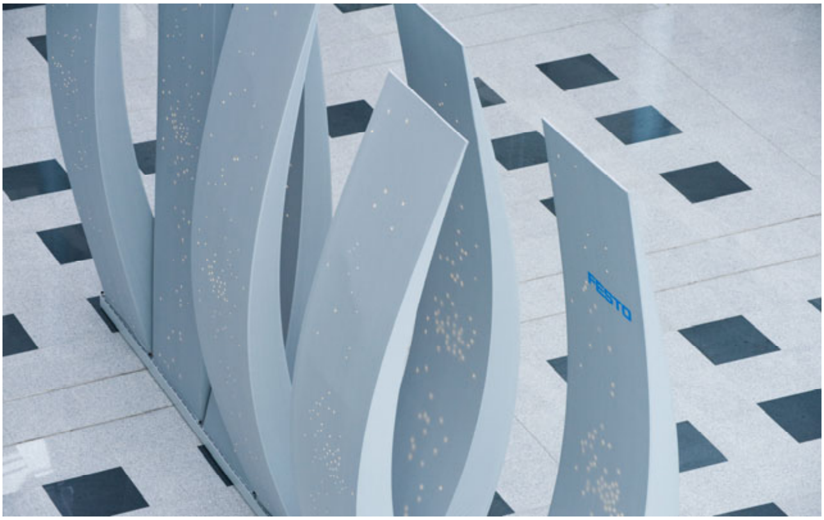


Fig. 4. Top: The InteractiveWall at the Hannover Messe 2009 displaying real time behavior by swinging its body back and forth, displaying patterns of light on its skin, and projecting localized sound.

4 Conclusion

An intuitive interaction, opinionated towards seamless information exchange is initiated through the research experiments, hence transforming everyday utilitarian space into an entertaining, inter-activating responsive organism. The prototypes fuel the idea of developing pro-active spaces communicating and reconfiguring in real-time, while being sensitive towards their context. The successful accomplishment of the projects is also suggestive of the benefits yielded by a collaborative manner of working with varying fields of expertise and stresses upon further envisioning emotive architectural beings which understand and respond to its occupants.

The aforementioned techniques of developing interdependent nodal networks, which double up as actuated details of interactive bodies, stress upon scripting localized relations between constituent details, thus actualizing the performance of the built form as an emergent, communicative exchange between the network of components. Information flow becomes a continual process in such real-time interactive prototypes, converting them into executable processing and reacting systemic entities.

Such architectural constructs eventually acquire the characteristics of living entities, sending and receiving information, processing this information locally, and producing optimal global output.

The intrinsic design decision of enriching the nature of architectural detailing and establishing an inter-disciplinary work process has significant impact on the nature of architectural space and its structuring principles. Such design-informatics-based hybrid typologies can be seen as highly logical systems, which, while being entertaining due to its shear grandeur and kinetic abilities, pave the path for performative responses to contemporary contextual complexities. These research experiments thus initiate an intuitive interaction, inclined towards seamless information exchange, transforming everyday utilitarian space into an inter-activating responsive experience.

References

1. Luhan, M.: Understanding Media. Routledge Classics, London (2002)
2. Lupton, E.: Skin, Surface Substance + Design. Laurence King Publishing Ltd., London (2002)
3. de Kerckhove, D.: The architecture of intelligence. Birkhauser, Basel (2001)
4. Srinivasan, A.V., Michael McFarland, D.: Smart structures analysis and design. Cambridge University Press, Cambridge (2001)
5. Reynolds, C.W.: Flocks herds and schools: A distributed behavioral model, Computer graphics. In: SIGGRAPH 1987, Conference Proceedings, vol. 21 (1987)
6. Bossomaier, T.R.J., Green, D.G.: Complex systems. Cambridge university Press, UK (2000)
7. Holland, J.H.: Emergence: from chaos to order. Oxford University Press, London (1998)
8. Kerkhove, D.: The architecture of intelligence. Birkhauser Publication, Basel (2001)
9. Oosterhuis, K.: Architecture goes wild. 010 publications, Rotterdam (2002)
10. Oosterhuis, K.: Hyperbodies: Towards an E-motive Architecture. Birkhauser, Basel (2003)
11. Aarts, E., Marzano, S. (eds.): The new everyday: views on ambient intelligence. 010 publishers, Rotterdam (2003)

Explorations in Player Motivations: Virtual Agents

Barbaros Bostan

Interactive and Digital Media Institute (IDMI), National University of Singapore
idmbb@nus.edu.sg

Abstract. Creating believable agents with personality is a popular research area in game studies but academic research in this area usually focuses on one facet of personality - for example, only on moods or character traits. The present study proposes a motivational framework to predict goal-directed behaviour of virtual agents in a computer game and explores the opportunities of using personality inventories based on the same motivational framework to design virtual agents with personality. This article claims that motivation to reach a goal is influenced by psychological needs which are represented with an equation that determines the strength of a character's motivational force. The framework represented by this study takes into account psychological needs and their interrelations for analyzing choices of virtual agents in a computer game.

Keywords: Personality, motivations, virtual agents, behavioural architecture.

1 Introduction

Creating autonomous believable virtual agents is one of the primary concerns of game industry but today's digital entertainment industry uses different approaches to establish virtual characters that behave the most believable way. Previous studies indicate that virtual agents are more believable if they have their own personalities, emotions, goals, and motivations; if they are perceptive of environmental changes; if they keep track of their own social interactions (memories); and if they are capable of reacting to player choices and actions throughout the game [1,2,3,4]. However, there is still no integrated or universally accepted model of personality that can be applied to virtual agents of a computer game. Academic research in this area usually focuses on one facet of personality - for example, only on emotions or character traits. In this regard, this article aims to analyze the goal-directed behaviors and personality aspects of virtual agents by using a motivational framework that takes into account psychological needs, emotional states, moods or feelings, general behavioral patterns and personality traits. Assuming that goal-directed behavior is triggered by the interaction between personal and environmental factors, the proposed motivational variables are derived from basic human needs. The needs investigated with this framework are psychological in nature, are not static entities but the result of both internal and external forces, and are concerned with mental and emotional states of a person.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 262–269, 2010.

2 A Need Framework for Virtual Agents

It is easy to understand the appeal of life-like virtual agents, but there is a great diversity of approaches on creating believable characters or characters with personality. To analyze the psychology of virtual characters, Paiva et al. [5] and Prada et al. [6] used emotional states and corresponding action tendencies; Lankoski et al. [7] used the the physiology-sociology-psychology three-dimensional structure with special emphasis on the protagonist of a story; Prada and Paiva [8] used the Five Factor Model of personality by focusing on extraversion and aggreeableness dimensions; Mosher & Magerko [9] used the Five-Factor Model of personality to define personality templates; Tychsen et al. [10] used the Personal Attributes Questionnaire - PAQ [11] to examine the relationship between players and characters; Rizzo [12] analyzed believability of virtual characters from an emotional state perspective; Ravaja et al. [13] examined emotional response patterns and sense of presence with the help of mood states and the impulsive sensation seeking dimension of Zuckerman-Kuhlman Personality Questionnaire - ZKPQ [14]; Kshirsagar and Thalmann [15] and Romano and Wong [16] used the OCC Model of Emotions [17] to represent the personality of social characters; and other researchers attempted different combinations of emotions, personality traits, moods, beliefs, and desires [18,19,20,21]. This article does not claim that these studies are irrelevant with the psychology of virtual agents; however, because each one of them focuses on different personality aspects the proposed models are not compatible with eachother.

The proposed personality model of this article is based on the needs framework of Murray [22], which defines psychological needs by matching them with actions and feelings, appropriate desires and effects, emotions, personality traits, actones (motones and verbones), different forms of need activity (intravertive, subjectified, semi-objectified, egocentric, infravertive, etc.), pathology, social forms and common relationships with other needs such as fusions, conflicts, and subsidiations. Although several classifications of needs exist in literature, this study regards each need as an individual variable of personality and uses twenty-seven needs of Murray's study: *Acquisition* (nAcq), *Construction* (nCons), *Order* (nOrd), *Retention* (nRet), *Aggression* (nAgg), *Blamavoidance* (nBlam), *Counteraction* (nCnt), *Defendance* (nDfd), *Deference* (nDef), *Dominance* (nDom), *Abasement* (nAba), *Affiliation* (nAff), *Nurturance* (nNur), *Rejection* (nRej), *Succorance* (nSuc), *Achievement* (nAch), *Autonomy* (nAuto), *Harmavoidance* (nHarm), *Infavoidance* (nInf), *Recognition* (nRec), *Exhibition* (nExh), *Cognizance* (nCog), *Exposition* (nExp), *Understanding* (nUnd), *Play* (nPlay), *Sentience* (nSen), *Sex* (nSex).

A recent study focused on matching twenty-seven needs of this framework to common gaming situations in computer role-playing games (CRPGs) [23]. Variables defined by major motivational studies in literature [24,25,26] are also matched by this study to corresponding psychological needs and/or components of goal-directed behaviour. In an attempt to take this study one step further and to identify the common interaction patterns between these individual needs, Bostan and Kaplancali [27] applied the same motivational framework to another computer game, analyzed psychological needs by defining the driving game mechanics behind them and identified common fusions of needs in the selected computer game (nAgg & nHarm, nRet & n Acq, nAcq & nDom, nCog & nDom). The same authors also analyzed user-created content (mods) of a popular computer game within the same motivational framework in terms of the needs they satisfy [28]. It is shown that, although the restrictions imposed by the game mechanics significantly reduce the number of player needs satisfied by a game and trap the player within the common motivational cycle of *Achievement*, *Aggression*, *Harmavoidance,* and *Acquisition* [27], but game mods created by users attempt to compensate for this by satisfying the needs of *Sentience*, *Exhibition*, *Recognition*, *Sex*, *Play* and *Affiliation* [28].

3 Predicting Virtual Agent Behaviour

Goal-directed behavior is made up of the interactions occurring between individuals and their environments that triggers actions and virtual character's computational processes are driven by their goals and preferences. There are situational and personal factors that influence the motivation to reach a goal. This article does not aim to identify all of the personal determinants of human behaviour and focuses only on personal factors. It is concerned with the possibility of using motives of a virtual agent to predict the likelihood of their behaviours. In this regard, for each gaming situation (GS), one or more needs will be dominant, which will be referred as Primary Needs (PN). Because some needs are in the service of others, whereas some others are opposing it, fusions of needs will be described with Secondary Needs (SN) and conflicting needs will be described with Opposing Needs (ON). For any gaming situation, the strength (M_f) of the motivational force of a behavior (B) is represented as M_f (B/GS) in Eq. 1. The strength of the motivational force is dependent on the personal factors (PN,SN,ON) that represent the personality of a virtual agent.

$$M_f(B/GS)= \sum PN + \sum SN - \sum ON \qquad (1)$$

The variables of Eq. 1 should be better understood with a concrete gaming example but first we need character profiles that compose of psychological needs. If rapid prototyping techniques can be used to create virtual agents, cost and production time reduction will be a major advantage. In this regard, although it is possible to use the original study of Murray, there are personality inventories in literature (Personality Research Form: PRF, Edwards Personality Preference Schedule: EPPS and Adjective Check List: ACL) that are based on psychological needs. For example, to broaden the scope of information about the nature of needs, Jackson's Personality Research Form [29] presents different correlates. Using these correlations, it is possible to build a quick character profile according to vocations, such as architect, chemist, and pharmacist; interpersonal behaviors, such as generosity, leadership, and objectivity; human values, such as aesthetic, economic, and religious; personality aspects, such as anxiety, responsibility, and innovation; vocational interests, such as medical security, business, and sales; or character traits, such as ambition, extraversion, and trustfulness. For example, it is possible to come up with a quick profile for a male dominant leader (A) or a character of trustfulness (B). According to Jackson's Personality Research Form, virtual agent A is characterized by high Dominance and Exhibition, and and low Sentience and Understanding. Similarly, virtual agent B is characterized by high Abasement, Affiliation, Nurturance, and Succorance and low Aggression, Autonomy, and Defendance. Sample profiles of these characters derived from [29] are given blow:

Table 1. Sample Character Profiles

	nAba	*nAch*	*nAff*	*nAgg*	*nAuto*	*nCog*	*nDfd*	*nDom*	*nExh*	*nHarm*	*nNur*	*nOrd*	*nPlay*	*nSen*	*nRec*	*nSuc*	*nUnd*
A	-12	12	15	10	-09	19	17	51	35	-02	02	10	-05	-31	14	19	-25
B	30	-13	31	-44	-27	05	-43	-16	-09	21	31	-08	17	-06	-06	29	-14

Assume that virtual agent A is the dominant military leader who faces a hostage situation, where a group of armed people took a number of hostages in a primary school. Not giving in to the demands of the villainous group, he orders the squad leader (virtual agent B) to raid the school and to eliminate any threats inside. When virtual agent B expresses his reluctance by stating that little children may get harmed in the way and one of the children is a friend's son, but virtual agent A threats virtual agent B with his career for disobedience to a direct command. In this gaming situation, A is using "autocratic power", a fusion of *Dominance* and *Aggression*, whereas B is driven by his need to protect children (*Nurturance*), especially his friend's son (*Affiliation*), but his career (*Achievement*) is in danger if he disobeys the command. For this GS, it is possible to define different primary and secondary needs for different situations, but we assume that the primary need of A is *Dominance* whereas *Aggression* is a secondary need and *Nurturance* is an opposing need. For virtual agent B, the PN is *Nurturance*, the SN is *Affiliation* and the ON is *Achievement*. Achievement is the opposing need of virtual agent B for this GS, but this character has a negative need for achievement which means that he does not care too much about his career. Thus, even if it is included in the equation as an ON, Achievement will behave like a SN for this GS. The motivational forces of each virtual agent are calculated below:

$$M_{fA}\,(B/GS)= \sum PN + \sum SN - \sum ON=51+10-2=59$$
$$M_{fB}\,(B/GS)= \sum PN + \sum SN - \sum ON=31+31-(-13)=75$$

The motivational forces of two virtual agents are opposing with each other but B's motivational force is stronger than A's. Thus, it is likely that B will disobey the command since the strength of the autocratic power employed by A is weaker than the combined motivational force of his psychological needs. But the important question is whether the secondary needs or the opposing needs are as important as the primary need. If the primary need is the dominant need for a psychological situation, the contributions of fusions and conflicts may be less influential on the total motivational force. A solution to this problem is to define the strength of the fusion or conflict, which will be represented with a multiplier of γ. Thus, for any gaming situation, the strength of the motivational force (M_f) of a behavior (B) will now be represented as M_f (B/GS) in Eq. 2.

$$M_f\ (B/GS) = \sum PN + \sum \gamma_{SN}.SN - \sum \gamma_{ON}.ON \qquad (2)$$

Assuming that the γ values have a default value of 0.5, the motivatonal forces of the two virtual agents for the GS in question are calculated again. Now B's motivational force is weaker than A's, which indicates an inclination towards obeying the order of virtual agent A.

M_{fA} (B/GS)= $\sum$ PN + $\sum \gamma_{SN}$.SN - $\sum \gamma_{ON}$.ON= 51+(0.5*10)-(0.5*2)=55
M_{fB} (B/GS)= $\sum$ PN + $\sum \gamma_{SN}$.SN - $\sum \gamma_{ON}$.ON=31+(0.5*31)-(0.5*-13)=53

The probability of the occurrence of a behavior can easily be calculated with the introduction of a random variable that represents the uncertainty in human behavior. Similar techniques can be found in table-top role playing games. For example, developed by Wizards of the Coast (formerly TSR), Advanced Dungeons and Dragons or AD&D is a game of imagination, as well as a game of tactics and strategy, which has set the standard for role-playing for more than 30 years. The accumulated knowledge of table-top role playing games has also influenced the design of popular CRPGs which use rules or game mechanics based on their pen-and-paper equivalents. Indeed, the majority of the most successful computer role playing games (CRPGs) such as *Baldur's Gate* series[1], *Neverwinter Nights* series[2], *Planescape Torment*[3], *Icewind Dale* series[4], *Knights of the Old Republic* series[5] and *Fallout*[6] series, use AD&D system or its variants. Understanding table-top role playing games and the mechanisms behind them can be useful in enhancing the quality of virtual agents. In this regard, these games have also drawn the attention of the researchers from the field of computer game studies [30,31,32]. These games represent randomness with dice rolls, for example 1d20 for a dice roll between 1 and 20. Thus, the modified version of Eq.2 with the addition of a random variable (R) is given below:

$$M_f\ (B/GS) = R + \sum PN + \sum \gamma_{SN}.SN - \sum \gamma_{ON}.ON \qquad (3)$$

For example, assuming that the R value will have a 1%-20% contribution to motivational force, if virtual agent A rolls a 10 on a 1d20 roll, the probability of the occurrence of a behavior can be calculated for virtual agent B. M_{fA} totals up to 65, and virtual agent B has to roll a 13 or better to oppose virtual agent B. So, there is only a 40% chance for virtual agent B to oppose the dominant leader's decision.

---

[1] Baldur's Gate (1998), Baldur's Gate II: Shadows of Amn (2000), both developed by Bioware.
[2] Neverwinter Nights (2002) developed by Bioware, Neverwinter Nights II (2006) developed by Obsidian Entertainment.
[3] Planescape Torment (1999) developed by Black Isle Studios.
[4] Icewind Dale (2000), Icewind Dale II (2002), both developed by Black Isle Studios.
[5] Knights of the Old Republic (2003) developed by Bioware, Knights of the Old Republic II: Sith Lords (2005) developed by Obsidian Entertainment.
[6] Fallout (1997), Fallout 2 (1998), both developed by Black Isle Studios, Fallout 3 (2008) developed by Bethesda Game Studios.

4 Conclusion

This study proposes a motivational framework for analyzing virtual agent behaviour in computer games; defines an equation to predict the likelihood of the occurence of a behavior in different gaming situations; and utilizes personality inventories to rapidly design characters with personality. This article is unique in that the proposed motivational framework does not focus on static facets of personality but takes into account psychological needs and interactions between these needs. For each need, using the original study of Murray [22], it is possible to make connections with emotions, feelings, desires, general behavioral patterns, and personality traits. Combining psychological variables/findings of other studies on psychological needs, such as the Jackson's Personality Research Form [29], this study provides an extensive knowledge base for character design and rapid prototyping of virtual agents.

This article does not claim to have identified all the factors that affect human behaviour, but it does try to address how the proposed psychological needs should facilitate the prediction of virtual agent behaviour in computer games. The next challenge is to integrate situational factors such as goal valence and user expectancy into the proposed equations but the biggest challenge in developing this framework is to relate these needs with trait models of personality in literature, such as the Big Five Personality Traits. Relating the proposed needs with other personality aspects, such as emotions which have drawn the attention of AI researchers [33,34], and coordinating them with lower level sets of actions will provide a unified model of personality. The proposed need framework provides a solid foundation for future research in computer games by providing methods for designing virtual agents and assessing the motivational aspects of non-player characters. If the behavioural equation and the character profiles can be validated by future empirical studies, the framework defined in this study should significantly improve our understanding of virtual agent behaviour in computer games.

References

1. Cutumisu, M., Szafron, D., Schaeffer, J., Mcnaughton, M., Roy, T., Onuczko, C., Carbonaro, M.: Generating ambient behaviors in computer role-playing games. IEEE Intelligent Systems 21(5), 19–27 (2006)
2. Livingstone, D.: Turing's test and believable AI in games. ACM Computers in Entertainment 4(1), 6–18 (2006)
3. Romano, D.M., Wong, A.K.L.: Personality model of a social character. In: 18th British HCI Group Annual Conference (2004)
4. Klıne, C., Blumberg, B.: The art and science of synthetic character design. In: Proceedings of the Symposium on AI and Creativity in Entertainment and Visual Art, AISB (1999)
5. Paiva, A., Dias, J., Sobral, D., Silva, C., Aylett, R.S., Woods, S., Zoll, C.: Caring for Agents and Agents that Care: Building Empathic Relations with Synthetic Agents. In: Proceedings of the Third International Joint Conference on Autonomous Agents and Multi Agent Systems, AAMAS 2004 (2004)
6. Prada, R., Vala, M., Paiva, A., Hook, K., Bullock, A.: FantasyA - The Duel of Emotions. Short paper at Intelligent Virtual Agents (IVA), Kloster Irsee, Germany (2003)

7. Lankoski, P., Heliö, S., Ekman, I.: Characters in Computer Games: Toward Understanding Interpretation and Design. In: Proceedings of Level Up: Digital Games Research Conference (2003)
8. Prada, R., Paiva, A.: Believable groups of synthetic characters. In: Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2005), pp. 37–43 (2005)
9. Mosher, R., Magerko, B.: Personality Templates and Social Hierarchies Using Stereotypes. In: Göbel, S., Malkewitz, R., Iurgel, I. (eds.) TIDSE 2006. LNCS, vol. 4326, pp. 207–218. Springer, Heidelberg (2006)
10. Tychsen, A., McIlwain, D., Brolund, T., Hitchens, M.: Player-Character Dynamics in Multi-Player Role Playing Games. In: Proceedings of DIGRA 2007, Tokyo, Japan (2007)
11. Spence, J., Helmreich, R., Holahan, C.: The Personal Attributes Questionnaire: A measure of sex role stereotypes and masculinity-feminity. Catalog of Selected Documents in Psychology 4, 43–44 (1974)
12. Rizzo, P.: Why Should Agents Be Emotional for Entertaining Users? A Critical Analysis. In: Paiva, A.C.R. (ed.) IWAI 1999. LNCS, vol. 1814, pp. 166–181. Springer, Heidelberg (2000)
13. Ravaja, N., Saari, T., Salminen, M., Laarni, J., Holopainen, J., Järvinen, A.: Emotional response patterns and sense of presence during video games: Potential criterion variables for game design. In: Proceedings of NordCHI 2004, Tampere, Finland (2004)
14. Zuckerman, M., Kuhlman, M., Joiremann, J., Teta, P., Kraft, M.: A Comparison of Three Structural Models for Personality: The Big Three, The Big Five, and the Alternative Five. Journal of Personality and Social Psychology 65, 757–768 (1993)
15. Kshirsagar, S., Thalmann, N.M.: Virtual Humans Personified. In: Proceedings of the first international joint conference on Autonomous Agents and Multi Agent Systems, AAMAS 2002 (2002)
16. Romano, D.M., Wong, A.K.L.: Personality Model of a Social Character. In: 18th British HCI Group Annual Conference, Leeds Metropolitan University, UK (2004)
17. Ortony, A., Clore, G., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge (1988)
18. Bailey, C., Katchabaw, M.: An emergent framework for realistic psychosocial behaviour in non player characters. In: Proceedings of the 2008 Conference on Future Play: Research, Play, Share, Ontario, Canada (2008)
19. Dugdale, J., El Jed, M., Pallamin, N., Pavard, B.: Modelling Character Emotion in an Interactive Virtual Environment. In: AISB Symposioum, Leeds, UK (2004)
20. Sandercock, J., Padgham, L., Zambetta, F.: Creating Adaptive and Individual Personalities in Many Characters Without Hand-Crafting Behaviors. In: Gratch, J., Young, M., Aylett, R.S., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 357–368. Springer, Heidelberg (2006)
21. Masuch, M., Hartmann, K., Schuster, G.: Emotional Agents for Interactive Environments. In: Proceedings of 4th Int. Conf. on Creating, Connecting and Collaborating through Computing, University of California, Berkeley (2006)
22. Murray, H.A.: Explorations in personality. Oxford University Press, New York (1938)
23. Bostan, B.: Player Motivations: A Psychological Perspective. ACM Computers in Entertainment 7(2) (2009)
24. Sweetser, P., Wyeth, P.: GameFlow: A model for evaluating player enjoyment in games. ACM Computers in Entertainment 3(3), 1–24 (2005)
25. Malone, T.W., Lepper, M.R.: Making learning fun: A taxonomy of intrinsic motivations for learning. In: Snow, R.E., Farr, M.J. (eds.) Aptitude, Learning, and instruction: Conative and Affective Process Analyses. Lawrence Erlbaum, Hillsdale (1987)

26. Yee, N.: Motivations of Play in Online Games. CyberPsychology & Behavior 9(6), 772–775 (2006)
27. Bostan, B., Kaplancali, U.: Explorations in Player Motivations: Game Mechanics. In: Proceedings of GAMEON 2009, Düsseldorf, Germany (2009)
28. Bostan, B., Kaplancali, U.: Explorations in Player Motivations: Game Mods. In: Proceedings of GAMEON-ASIA 2010, Shanghai, China (2010)
29. Jackson, D.N.: Personality Research Form Manual. Research Psychologists Press, Port Huron (1989)
30. Tychsen, A., Hitchens, M., Brolund, T., Kavakli, M.: The Game Master. In: Proceedings of the Interactive Entertainment Conference, Sydney, Australia, pp. 215–222 (2005)
31. Peinado, F., Gervas, P.: Transferring game mastering laws to interactive digital storytelling. In: Göbel, S., Spierling, U., Hoffmann, A., Iurgel, I., Schneider, O., Dechau, J., Feix, A. (eds.) TIDSE 2004. LNCS, vol. 3105, pp. 1–12. Springer, Heidelberg (2004)
32. Louchart, S., Aylett, R.: Towards a narrative theory of virtual reality. Virtual Reality 7(1), 2–9 (2003)
33. Bosser, A., Levieux, G., Sehaba, K., Buendia, A., Corruble, V., Fondaumiere, G.D.: Dialogs taking into account experience, emotions and personality. In: DIMEA, pp. 9–12 (2007)
34. Sehaba, K., Sabouret, N., Corruble, V.: An Emotional Model for Synthetic Characters with Personality. In: Proceedings of the 2nd international Conference on Affective Computing and intelligent interaction, Lisbon, Portugal (2007)

Integration of CityGML and Collada for High-Quality Geographic Data Visualization on the PC and Xbox 360

Marc Herrlich, Henrik Holle, and Rainer Malaka

Digital Media Group
TZI, University of Bremen
Germany
mh@tzi.de, henrik@holle.net, malaka@informatik.uni-bremen.de

Abstract. Computer games and serious geographic information systems (GIS) share many requirements with regard to storage, exchange, and visualization of geographic data. Furthermore, there is a demand for high-fidelity photo-realistic and non-photo-realistic visualization. This poses at least two questions: Is there a single data format standard suitable for serious GIS-based applications and computer games that supports state-of-the-art visual quality? How can computer games and serious applications benefit from each other, especially platform-wise? In this paper we will investigate both questions by taking a closer look at the CityGML standard in comparison to COLLADA and we will report on our findings in integrating CityGML with mainstream game technology. The main contribution of this paper to the field is a suggested way of integrating an important features of CityGML and Collada for high-quality visualization, i.e. programmable shader effects, and demonstrating the feasibility of employing a game console as a cheap and widely available device for geodata visualization and possibly other geodata-centric applications.

1 Introduction

Currently, even in areas of mainstream information technology the terms *Serious Games* and *Serious Gaming* seem to be ubiquitous, albeit often used in a somewhat general manner. Even so, this development shows that the industry and the scientific community alike have come to recognize the benefits of employing game technology for other purposes. Besides research and applications directly targeted at exploiting the motivational and educational aspects of games, the scientific community values the comparatively cheap and powerful platform for mainstream hardware provided by game technology and the industry clearly appreciates the cost benefits, especially in areas where dedicated high-performance solutions are/were predominant, e.g. in military simulation. While there is an ongoing discussion among practitioners and theorists if the term Serious Games really applies to the general usage of game technology for non-entertainment applications that have no or very limited game-like appeal, this question is only of

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 270–277, 2010.

limited importance for the scope of this paper and therefore we will not directly address it. For reference, we point to the taxonomy of Serious Games by Sawyer [1][1]. In our opinion the "seriousness" and/or "gameiness" of any game or application very much depends on the context and intention of usage and perception of the user.

Games as applications tend to quickly follow and adapt developments from other areas, e.g. trends like user-generated content or the interest in geographic information systems (GIS), sparked by applications such as Google Earth[2] or Microsoft's Virtual Earth[3]. As game technology is of rising relevance to many other areas of information technology, it is equally important to investigate how both sides can benefit most from each other and what standards and techniques are available or needed. In this paper we will concentrate on the area of GIS-based applications. Serious GIS-based applications and games both have a high demand for *direct* geodata, i.e. geodata resulting from measurements and suitable post-processing, and *indirect* geodata, i.e. geodata that is the result of a modeling process that might or might not have any relations to the real world. Although, in serious applications most often direct data is dominant and in games indirect data is prevailing, it is often the case, that games also include direct data, e.g. in sport, racing, or simulation games, and that serious applications make use of indirect data, e.g. Google Earth. Therefore, it is consequential to take a closer look at the requirements of both applications and discuss the standards and techniques currently used or proposed. For the scope of this paper we will focus on discussing the possibilities, standards, and techniques of using direct geodata for games and applications based on game technology and show how to easily integrate the foundation for state-of-the-art rendering, i.e. so-called "shader effects", with CityGML as a geodata standard. Furthermore, our prototype also runs on the Xbox 360, a so-called "next-gen" gaming console with a large user-base.

The rest of this paper is organized as follows. In section 2 we present some related works. In section 3 we discuss essential features of two important standards for geodata, i.e. CityGML and Collada, with regard to the requirements of applications concerned with high fidelity visualization. In the same section we suggest a small but important extension for CityGML in order to remedy one of the major shortcomings we discovered in our experiments. A prototype we created based on the XNA framework is presented in section 4. In section 5 we summarize and reflect our findings and conclude the paper by outlining open questions and future research directions.

2 Related Work

Already in 2002, Rhyne [2] argues that scientists today had much to learn from the computer games industry regarding computer graphics, visualization, and

[1] Conference presentation available online at `www.seriousgames.org`

[2] `earth.google.de`

[3] `www.microsoft.com/virtualearth`

interfaces. In Rhyne's opinion, scientific visualization applications could benefit from computer game technology, as most computer games are optimized for commodity hardware. She also states some drawbacks in using game engines with respect to data accuracy and reliability. Since 2002 computer games have surely continued to push the state-of-the-art in computer graphics and additionaly we think that many scientific applications could also benefit from many of the interaction metaphors used in computer games, although this is not the focus in this paper. Nearly all computer games with a rich graphical world rely on so-called shader effects, which are "short" programs most often written in a special C-like language. We show in this paper how to include such programs, which are essential for high-quality real-time graphics today, into semantically rich data standards for geographic data.

Herwig and Paar [3] discuss the suitability of game engines for landscape visualization and planning. They present different usage scenarios and analyze the requirements of landscape architects concerning supporting tools and to what extend game engines can solve these problems. Fritsch and Kada [4] discuss indoor as well as outdoor visualization of geodata based on different game engines. They also compare the benefits of game engines in contrast to other software libraries and present concepts for integrating them with other tool for purposes like computer aided facility management. They conclude, that the conversion process of geodata into the data format of a game engine was one of the major obstacles for such applications. Both the works of Herwig and Paar and of Fritsch and Kada clearly demonstrate the suitability of computer game technology for general visualization and especially for geodata visualization applications. However, while theirs and some similar works deal with some of the fundamentals, they do not concern themselves with integrating essential features directly at the data format level, nor do they include game consoles as a platform.

Zeile et al. [5] present a case study of creating and visualizing a virtual 3d city model of Bamberg. Their report clearly demonstrates the complexity of integrating the huge number of different data sources. They also state that "aesthetic ambitions" and "degree of detailing" had been important criterions. This supports our assumption of a demand for high-quality real-time graphics in this application area, which is the domain of computer games employing shader programs.

Döllner et al. report in [6] and [7] on the backgrounds of their LandXplorer system and their usage of CityGML. In the presented general system architecture, game engines and our work would fit in well as "3D City Model Presentation Systems" component.

Kibria et al. discuss in [8] the requirements of virtual environments based on geodata. They suggest, that visual fidelity and - more general - realistic surface materials were very important for the users of such environments, which might require the additions to geodata standards we discuss in sections 3 and 4 of this paper.

3 Geodata Standards

In recent years, the demand for three dimensional city models has increased considerably. More and more data of cities, regions, and states is captured and evaluated. Based on this data, three dimensional city models are created and procured. Out of this development process many standards for geodata storage, exchange, and visualization have evolved. Some standards such as the Geography Markup Language [9] [10], Keyhole Markup Language [11], digital ground models (DGM), or ESRI Shapefiles [12] focus on the exchange of geometric data with geographic references. Other standards like X3D [13] or Collaborative Design Activity (COLLADA) [14] have been developed for the exchange of generic 3d scenery. In addition to topological and geometrical data, city models include a semantic model. The common data exchange standards like GML, X3D or COLLADA do not feature this. The comparatively new format CityGML [15] targets the problem by unifying geometric data and semantics into one data exchange format. Therefore, CityGML provides a decent format for the efficient and lossless data exchange within geodata content pipelines and is increasingly used by (German) land surveying offices. The use of the CityGML format enables a high degree of semantic and syntactic interoperability between applications. However, while CityGML models might include very rich semantic information, it is not clear if CityGML provides sufficient support for applications with high demands for photo-realistic visualization. The following paragraphs analyze the format regarding this aspect of CityGML, i.e. as a basis for applications generating high fidelity visualization.

3.1 CityGML

The Open Geospatial Consortium Inc. (OGC) has adopted CityGML 1.0.0, the open data model developed for the storage and exchange of virtual 3d city models, as an official standard. It is implemented as an Geography Markup Language (GML) application schema and uses only a subset of the complex GML profile. Unlike formats such as X3D or COLLADA, CityGML does not only represent the graphical appearance of city models but allows the modeling and representation of the semantic and thematic properties, respectively, i.e. taxonomies and aggregations of digital ground models, buildings, vegetation, water bodies, transportation facilities, and city furniture. The integration of geometric and thematic data into one format allows to utilize 3d city models for more use cases aside from pure visualization. The CityGML model distinguishes between four different level-of-details, from a simple block model (derived by extruding ground planes), up to architectural building models (out- and indoor).

3.2 Extending CityGML

As mentioned before, serious GIS-based applications and games both have a high demand for *direct* geodata. CityGML, the 3d city model exchange format, could be the data source for *direct* geodata for visualization purposes. The following paragraph compares the geodata exchange format to the general and widely used 3d scenery exchange format COLLADA regarding its applicability to visualization purposes. On closer inspection, the CityGML data model can be divided into three different types of data: geometrical and thematical information and data concerning the appearance of objects.

CityGML's geometric model is derived from GML and mainly uses the basic type "polygon", which can be combined to complex surfaces. A thematic feature type, e.g. roof or door, can be assigned to each surface type. The modeling approach follows the boundary representation concept [16]. COLLADA's geometric modeling approach also follows the boundary representation concept but additionally supports polygonal meshes and modeling techniques closely related to modern graphic hardware programming interfaces.

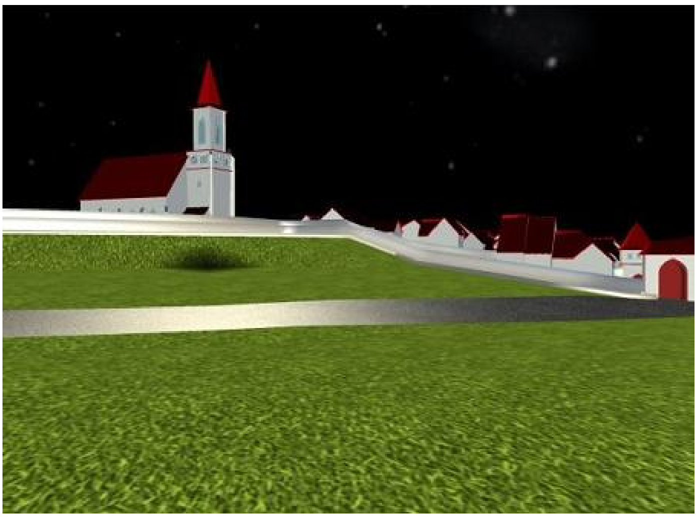


Fig. 1. Prototypical textures derived from the thematical model

The thematical model is a unique feature in CityGML. COLLADA and other formats for the exchange of general 3d scenery do not feature this. They exclusively offer the modeling of geometrical data and surface properties, i.e. textures, materials, and (hardware) shader programs. Applications focusing on visualization have no direct need for the thematical model. Nevertheless, if thematical data is available, it can be used to enrich the presentation in case the appearance model, i.e. surface data, is missing or incomplete. Colors or prototypical textures could be derived by feature type (figure 1), e.g. prototypical textures for land use and roads, or coloration of roof and wall feature types. Instead of using photo-realistic rendering, non photo-realistic rendering techniques [17] can also be applied.

An area of concern regarding CityGML is the appearance model. Compared to general 3d scenery exchange formats, CityGML's appearance model just allows minimal modeling of presentation parameters. The current CityGML version only supports three different types of texture mapping and a very basic lighting and material model. COLLADA for example allows the definition of complex materials including shader programs or shader fragments using its COLLADA FX extension. For high-fidelity photo-realistic visualizations the use of shading techniques like bump or normal mapping are common. The limitation can be solved by combining COLLADA FX and CityGML using CityGML's application domain extension (ADE) functionality. The COLLADA FX material library can be embedded into a CityGML document. A new appearance type derived from the original appearance type can reference the COLLADA FX material. This extension allows backwards compatibility to existing geodata infrastructures on the one hand, and allows the definition and lossless exchange of complex material definitions on the other hand.

The implementation details of a CityGML ADE for ColladaFX are quite straight forward (albeit tedious because of the lengthy ColladaFX specification). Basically, the respective ColladaFX schema definitions will have to be imported into a common schema definition file. Assuming such a common schema definition, Collada materials based on shaders effects could be used by defining a special CityGML surface type.This surface type could then be assigned to CityGML features that shall be rendered with shader effects.

4 Geodata Visualization on the Xbox 360

We created a sample implementation for the visualization of CityGML models using the Microsoft XNA framework. The XNA framework is widely used by hobbyists and even an increasing number of professionals to create games for the PC but also for the Xbox 360 gaming console. Furthermore, it is more and more used as a managed graphics API for generic applications, including scientific visualization and other serious applications. On important feature of the XNA framework is the so-called "Content Pipeline", which provides a clean interface and workflow for using all kinds of assets, i.e. 3d models, textures, sounds, and more.

We extended the XNA Content Pipeline[4] to enable the easy import and conversion of CityGML files. XNA content importers and processors for CityGML have been implemented. The importer loads the CityGML file, validates it against the schema specification and transforms it into a more general XML DOM representation. The processor transforms the CityGML model into an optimized format for the XNA graphics interface and performs important steps, like tesselating polygonal data to primitives stored in vertex and index buffers, generating texture coordinates if needed, creating/loading shader effects for lighting and texturing, and partitioning the data using common data structures like quad trees. The resulting data can be read and used by any XNA application. Internally, the CityGML processor is divided into several sub-processors for specific categories of CityGML objects, e.g. buildings, water, or relief features.

[4] msdn.microsoft.com/en-us/xna/default.aspx

```
//select all top level featuretypes via xpath
XmlNodeList featureTypeNodes =
  xmlDOM.SelectNodes("CityModel/cityObjectMember");

foreach(XmlNode xmlNode in featureTypeNodes)
{
  //Create processor via factory pattern
  CityObjectProcessor processor =
    CityObjectProcessorFactory.MakeProcessor(xmlNode);
  //set processing content node to current featuretype
  processingContext.Node = xmlNode.LastChild;
  //start processing
  processor.Process(processingContext);
}
```

Fig. 2. XNA CityGML Processor

A factory pattern is used to create the matching processor directly from the XML DOM representation using XPATH queries for retrieving top level features (figure 2).

5 Discussion and Future Work

Because our sample utilizes the XNA Content Pipeline, it runs not only on the PC platform but also enables the use of CityGML based applications on the XBox360 platform.

Compared to common 3d scenery exchange formats, CityGML is not designed for high fidelity visualization applications due its rudimentary presentation model. The integration of COLLADA FX into CityGML as an application domain extension solves this issue and combines the advantages of both standards. The introduced XNA integration verifies on the one hand that CityGML data can easily be transformed to a real-time rendering friendly format using the XNA Content Pipeline and on the other hand that mainstream game technology is well suited as basis for serious GIS applications. There are many possible directions for future work. The questions remains, how to utilize the (theoretically) rich semantic information for entertainment applications, e.g. by enriching the game play based on the semantics of objects/areas in the world. Furthermore, it could be useful to integrate additional features from the COLLADA side, like COLLADA objects or physics. We would also like investigate further about the acceptance/demand and feasibility of serious GIS-based applications on the Xbox360 and other game platforms.

References

1. Sawyer, B., Smith, P.: Serious games taxonomy. Presentation at the The Serious Games Summit at the Game Developers Conference (2008)
2. Rhyne, T.M.: Computer games and scientific visualization. Commun. ACM 45, 40–44 (2002)

3. Herwig, A., Paar, P.: Game engines: Tools for landscape visualization and planning? In: Buhmann, E., Nothhelfer, U., Pietsch, M. (eds.) Trends in GIS and Virtualization in Environmental Planning and Design. Wichmann, Heidelberg (2002)
4. Fritsch, D., Kada, M.: Visualisation using game engines. Archiwum ISPRS 35 (2004)
5. Zeile, P., Schildwächter, R., Poesch, T., Wettels, P.: Production of virtual 3d city models from geodata and visualization with 3d-game engines. a case study from the unesco world heritage city of bamberg. In: Buhmann, P., Bishop, L. (eds.) Trends in Real-Time Landscape Visualization and Participation, Wichmann, Heidelberg (2005)
6. Döllner, J., Baumann, K., Buchholz, H.: Virtual 3d city models as foundation of complex urban information spaces. In: Proceedings of CORP 2006 & Geo-Multimedia (2006)
7. Döllner, J., Kolbe, T.H., Liecke, F., Sgouros, T., Teichmann, K.: The virtual 3d city model of berlin - managing, integrating, and communicating complex urban information. In: Proceedings of the 25th Urban Data Management Symposium UDMS (2006)
8. Kibria, M.S., Zlatanova, S., Itard, L., van Dorst, M.: GeoVEs as Tools to Communicate in Urban Projects: Requirements for Functionality and Visualization. In: 3D Geo-Information Sciences, pp. 379–395. Springer, Heidelberg (2008)
9. Open Geospatial Consortium Inc.: OpenGIS Geography Markup Language (GML) Encoding Standard (2007) (White Paper)
10. Kolbe, T.H.: Representing and exchanging 3d city models with citygml. Lecture Notes in Geoinformation and Cartography, pp. 15–31. Springer, Heidelberg (2008)
11. Open Geospatial Consortium Inc.: OGC KML (2008) (White Paper)
12. Environmental Systems Research Institute: ESRI Shapefile Technical Description (1998) (White Paper)
13. ISO: ISO/IEC 19775:2004 – Information technology – Computer graphics and image processing – Extensible 3D (2005) (White Paper)
14. Khronos Group: COLLADA 1.5.0 Specification (2008) (White Paper)
15. Open Geospatial Consortium Inc.: City Geography Markup Language (CityGML) Encoding Standard (2008) (White Paper)
16. Foley, J.D., VanDam, A., Feiner, S.K.: Computer Graphics: Principles and Practice in C. Addison-Wesley Longman, Amsterdam (October 1995)
17. Lansdown, J., Schofield, S.: Expressive rendering: a review of nonphotorealistic techniques. Computer Graphics and Applications, IEEE 15, 29–37 (1995)

Virtual Blowgun System for Breathing Movement Exercise

Peichao Yu, Kazuhito Shiratori, and Jun'ichi Hoshino

University of Tsukuba, Graduate school of Systems and Information Engineering,
1-1-1 Tennodai, Tsukuba-shi, Ibaraki, Japan
admin@nerocrux.org, c_tori@mbb.nifty.ne.jp,
jhoshino@esys.tsukuba.ac.jp

Abstract. Breathing is the most basic requirement for having good health. However, unhealthy breathing like overbreathing and hyperventilation will happened easily without any awareness. We propose an experimental breathing movement exercise system - Virtual Blowgun System (VBS), offering an easy way of breathing exercise, for people of different physical strength, without space and safety limitations.

Keywords: breathing movement exercise, view-dependent virtual reality, versatile training field, blowgun interface.

1 Introduction

Breathing is the most basic requirement for having good health. However, unhealthy breathing like overbreathing and hyperventilation will happened easily without any awareness. It upsets the basic rhythm of life and the body works less effectively by causing blocked nose, insomnia, snoring, sleep apnea, panic, anxiety and depression. [1] Furthermore, unhealthy breathing may also cause much more serious diseases, such as allergies, asthma, COPD and emphysema.

There are several breathing methods, such as Buteyko or Pranayama [2], proved to be effective in the treatment of diseases in breath, or in breathing enhancement.

In Japan, a more popular breathing method called "abdominal breathing method" is easy to practice, especially through playing a traditional Japanese sports game, sports blowgun. Through playing this game in a proper way, breathing can be well controlled and exercised. [9]

However, sports blowgun requires wide space and strict safety care. Children, elderly people, or people with disabilities can hardly playing this game, and will not able to compete with ordinary mature adults.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 278–285, 2010.

We propose an experimental breathing movement exercise system - Virtual Blowgun System (VBS). The aim of VBS project is to offer an easy way of breathing exercise, for people of different physical strength, without space, safety or other limitations.

VBS will offer a highly simulated physical and visual experiment as playing the real blowgun sports. Users will also enjoy their exercises with game contents designed for breathing exercise.

2 Relevant Study

Wijnand's research [3] shows the importance of virtual environment on helping boost motivation for users training at home.

Virku[4] makes it possible for users to control the avatar in the game by using the exercise bicycle. Game contents are used to help keeping motivation of training.

Tsuji's research [5] on sports skill training focuses on the controlling of the movement of users' body parts.

Masuko's research [6] reflects heart rate on exercise game, which provides an easier way of designing applications, through which both user's motivation and exercise efficiency can be proved.

These systems are able to help user training at home, using virtual reality. However, we found out that the interactive is quite simple. Visually, users can hardly acquire immersive feelings, which will attenuate their motivation of training.

There are devices designed for breathing exercise or long-term medical treatment [7], but there are no interaction, no competition when using these devices. Without these elements, almost all users will feel tired and boring and lack of motivation in a couple of hours after using these devices.

Through our research, we found out that it is important and necessary, to enable virtual reality in creating training system, in order to provide interaction and challenge of archieving goals, which provides motivation of exercising.

3 System Concept and Design

VBS consists with 3 parts: blowgun device, displays and USB camera. A PC is also required for data processing and 3D scene generation.

We use large-scale design, for using in exhibitions and public facilities, which requires 2 projectors for front projection, and another projector for floor projection. System setup plan is shown on *Fig. 1*.

The system can also be used at home or office, by simply connect the blowgun device to the PC through USB interface. With a USB camera connected, users can also enjoy breathing exercises, sitting or standing in front of a personal computer, by playing game contents for breathing exercises.

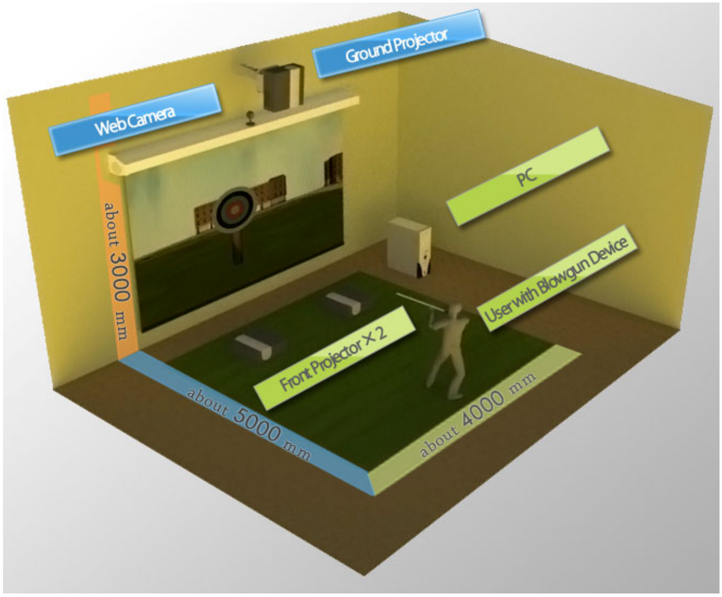


Fig. 1. Floor plan of VBS

Fig. 2 shows the specific structure and dataflow of VBS, which consisted with three parts: virtual blowgun device, USB camera for View-dependent Virtual Reality and PC application.

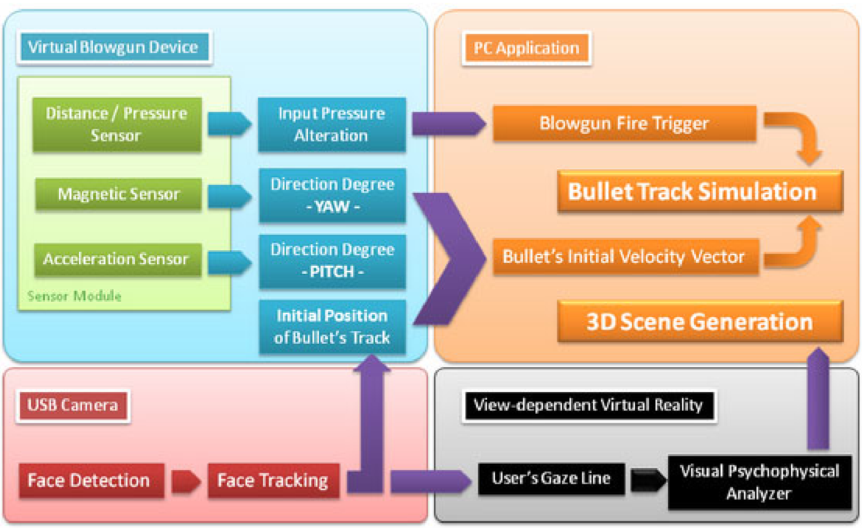


Fig. 2. System structure of VBS

3.1 Virtual Blowgun Device

The pipe-like device is the main input device of our system. We use a aluminum pipe (length of 1200mm for training field and exhibition use, 500mm for home use, width of φ15mm) as the body of the device.

On one side of the pipe, sensor module and pressure controller are installed. And on the other side of the pipe is replaceable mouth piece, through which user blow air into the device.

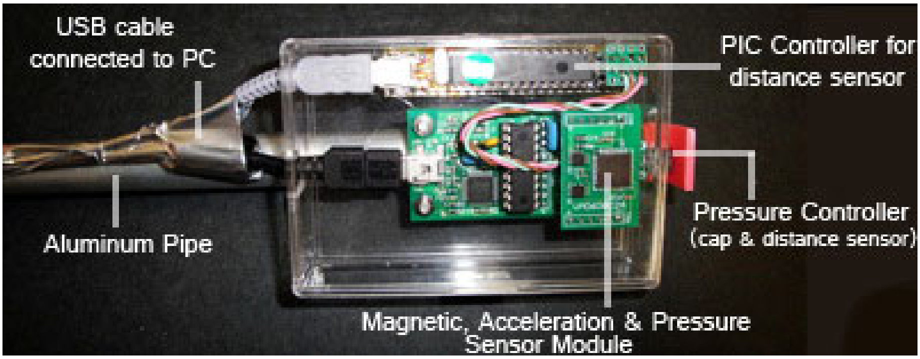


Fig. 3. VBS device structure

Fig. 3 shows the design of the key part: the sensor module and pressure controller of the VBS device. Currently we are still working on the prototype of the device, so all parts of the device have not been combined into one board yet. We have a PIC controller for data processing, and an IR distance sensor, a magnetic sensor, an acceleration sensor and a pressure sensor installed on the pipe.

A pressure controller is also designed, for measuring and controlling of the pressure inside the pipe.

3.2 USB Camera

USB Camera is designed to be set upon the center of the display. Manual calibration (inclination angle, position) is required to make sure user's face is in the center of the image camera captures.

The image taken by the camera is used for the generation of view-dependent virtual reality. Details will be explained in Section 4.1.

3.3 PC Application and Displays

PC application we created does the following works:

- Processes all the data retrieved from all sensors through serial ports.
- Graphical representation of 3D scene using OpenGL.
- Face detection and tracking using OpenCV.
- Audio effects using OpenAL.

The virtual scene we built is based on the image of European old town. We suppose the training is on the grass outdoor. The front and ground projectors working seamlessly, offer users a half-surrounded 3 dimensional VR exercise environment.

3.4 Concept and Rules

The user is asked to hold the VBS device and try his/her best to aim the center of the target and blow into the device.

When the pressure inside the device touches the trigger point, a virtual bullet will be generated and shown on the screen, which has the same physical behavior as the real bullets do. This will be explained in Section 4.2. The application will detect the collision of the bullet and the target, calculate the point user get.

We designed the VBS system as a virtual training field now, which uses big-size screens. However, the system can also be used for home or office, by simply connect the VBS device (USB) and USB camera to PC.

4 Visual Simulation

4.1 View-Dependent Virtual Reality Using Face Tracking

We propose a simple way of increasing the quality of virtual environment visually, using face tracking technologies, which requires no markers or devices binding on user's body. Only a USB camera, set on the top of the display device, and appropriate lightening are required.

We are currently using faceAPI from seeingmachines[8] for 3D face tracking. Both face's position and orientation is tracked real time, in order to measure raw gaze line.

Through our research we found out that instead of moving eyeball, user prefers move head when he is trying to gaze on something during exercise. So we just simply assume that there is no eyeball movement and the normal vector of user's face can be considered as user's gaze line.

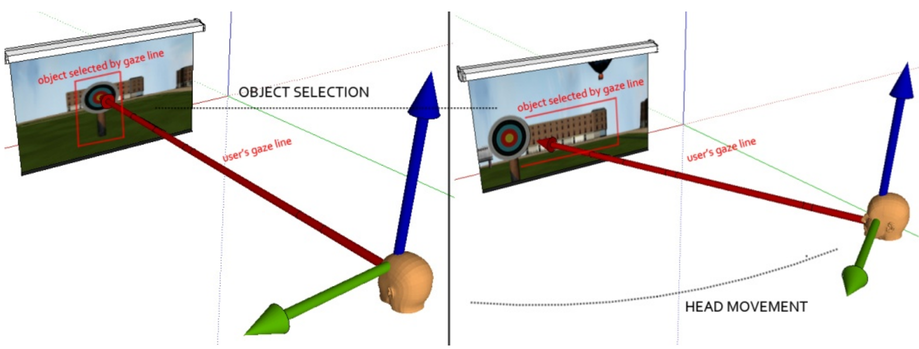


Fig. 4. Synchronize of user's face and virtual camera, and gaze line-base object selection

We suppose user's initial view point is to the center of the display, and user's face is also in the center of the image captured by the camera. When user's head moves, the virtual camera in the application will also moves, synchronized with user's gaze line. (*Fig. 4*)

Not only the VR camera's position, but also the position the camera focuses on will be changed. Consider the target is on the center of the screen, when user is trying to aim, or focus his sight on the target, he/she will probably move his/her face above the center point. When user moves his/her face away from center, he/she is probably trying to see something behind the target or something far away.

We consider the effect of visual psychophysics, by using deformation, zoom-in, blurring, etc, to enhance user's visual experience.

User's gaze line is supposed as a beam, shot into the virtual scene generated by the application, and hits the object in the scene. The gaze line can be used in several ways:

- The angle between the gaze line and the object's front plane normal can be calculated so we can create deformation on objects in the VR scene, raise the quality of immersive feelings. [10]
- When user gazing at an object for a period of time, trying to focusing on something (take aim at something for example), objects irrelevant to this action will be blurred.

4.2 Physical Simulation of Blowgun Bullet

We integrate the virtual training space into real world. One unit in the virtual space is set as 1/1000 meter. As shown in *Fig.5*, in real world, distance from user to screen is 3 meters, in virtual space distance from screen to target is 7 meters. Height of target will be set the same as user's height, so different people at different height will not feel uncomfortable when using the system.

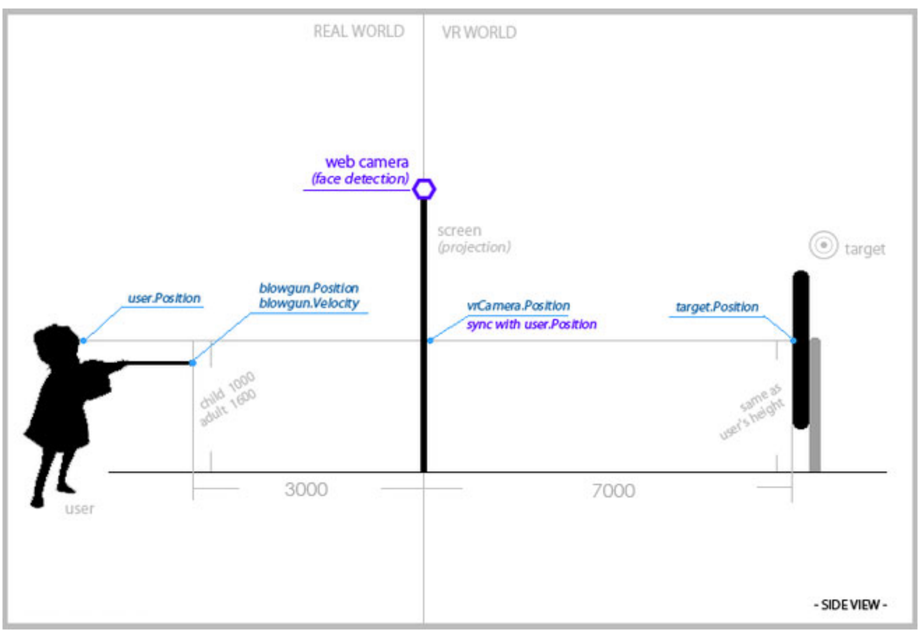


Fig. 5. Integration of virtual world with real world

We have built a physical model, which can simulate the movement of the bullet, considering the effect of the gravitation and the resistance forces.

The simulation depends on two parameters: the initial velocity vector and the initial XYZ position of the bullet.

Initial velocity vector can be acquired from sensor module. The 3-axis magnetic sensor for YAW-angle, and the 3-axis acceleration sensor for PITCH-angle.

Concerning magnetic field changes due to location or the interference of electronic and magnetic devices like PC, magnetic sensor needs to be initialized, by simply point the VBS device vertically to the display. We also need to know the scalar initial speed of the bullet, which we set to 50m/s. This value will be associated with the inner pressure of the device in our future work.

Initial XYZ position of the bullet is considered to be at the same position of user's face.

The application will check the collision between the bullet and the target, and will be marked clearly to tell user whether the bullet is hit the target or not.

5 Future Works

Our research will be continued and focused on the following two major themes:

1. Method of creating highly simulated versatile training field, which can also be easily setup at home or office with rare limitations.
2. The application of view-dependent virtual reality with face tracking technologies in versatile training fields.

Currently, the design of the VBS device is still in experimental stage. The following parts of the device will be re-designed or added:

- Automatic pressure control unit.
- Cable-free VBS device. We will use wireless module (bluetooth) instead of USB module for sensors' data transferring.
- The sensor unit will be merged into one narrow board, and as small as possible.

We also have plans of:

- Creating game contents which are able to control user's breathing movement.
- Enable online battles between players.

References

1. `http://www.buteykolearning.com/`
2. Cooper, S., Oborne, J., Newton, S., Harrison, V., Thompson Coon, J., Lewis, S., Tattersfield, A.: Effect of two breathing exercises (Buteyko and pranayama) in asthma: a randomised controlled trial. Thorax Aug. 58(8), 674–679 (2003)
3. IJsselsteijn, W., de Kort, Y., Westerink, J., de Jager, M., Bonants, R.: Fun and sports: Enhancing the Home Fitness Experience. In: Rauterberg, M. (ed.) ICEC 2004. LNCS, vol. 3166, pp. 46–56. Springer, Heidelberg (2004)

4. Mokka, S., Väätänen, A., Heinilä, J., Välkkynen, P.: Fitness Computer Game with a Bodily User Interface. In: Proceeding of the Second International Conference on Entertainment Computing, pp. 1–3 (2003)
5. Tsuji, T., Sumida, Y., Kaneko, M., Sadao: A Virtual Sports System for Skill Training. Journal of Robotics and Mechatronics 13(2), 168–175 (2001)
6. Masuko, S., Hoshino, J.: An Exercise Game Reflecting Heart Rate. The Journal of the Society for Art and Science 6(3), 136–144 (2007) (in Japanese)
7. Yuuichirou, M., Ishiguro, K., Yasuo, O.: Simplified Apparatus for the Abdominal Breathing Exercise. The Journal of Japanese Physical Therapy Association 22(7), 413–416 (1995) (in Japanese)
8. `http://www.seeingmachines.com/product/faceapi/`
9. Takayuki, N., Terukazu, H., Yokiko, G., Yukiko, M., Hirotaka, O., Hironobu, Y., Hideaki, F., Satoru, K., Seiichiro, T.: Effects of Blowgun Training on Respiatory Functions: For Applied Pulmonary Therapy. The Journal of Kyushu University of Nursing and Social Welfare 9(1), 3–6 (2007) (in Japanese)
10. Owada, S., Fujiki, J.: Ambiguous Rendering for Suppressing Deformation Effect for Stereoscopic Viewing of Spatially Inconsistent Scenes. ITE Technical Report 32(27), 21–24 (2008) (in japenese)

Development of a Virtual Electric Wheelchair – Simulation and Assessment of Physical Fidelity Using the Unreal Engine 3

Marc Herrlich[1], Ronald Meyer[1], Rainer Malaka[1], and Helmut Heck[2]

[1] Digital Media Group
TZI, University of Bremen
[2] Research Institute Technology and Disability (FTB)
of Evangelische Stiftung Volmarstein

Abstract. This paper demonstrates how an existing game technology as a component off-the-shelf can be used as a basis to build a serious game for assistive technology for disabled people.Using the example of a virtual electric wheelchair simulator, we present how to use a computer game physics engine to achieve a realistic simulation of driving an electric wheelchair in a virtual environment. Focus of the paper is the conversion of driving characteristics of prevalently used electric wheelchairs into the virtual physics system of the used computer game engine. The used parameters are systematically balanced between the virtual and the real world to evaluate the realism of the driving characteristics of an electric wheelchair using the integrated physics simulation of the Unreal Engine 3.

Keywords: serious games, virtual electric wheelchair, COTS, physics simulation.

1 Introduction

Users of wheelchairs quite often face limitations and barriers when moving in public and private environments due to the fact that many environments have not been designed to be accessible for wheelchairs [8]. Awareness and knowledge of the needs of disabled people needs to be increased, especially for people who provide public services and for people who design public environments. A serious game [6] can be a good tool to rise awareness. For having a connection between the physical and the virtual world a certain degree of realism in the virtual world is necessary, meaning in our case the physical driving behavior of the wheelchair should be approximated as good as possible. The Unreal Engine 3 (UE3), a state of the art game engine, includes the Nvidia PhysX physics engine, and thereby offers the functionality to realize a virtual driving simulator for a wheelchair.

The quality of the approach to simulate the specific driving characteristics is evaluated in this paper including the application of the working wheelchair model in a virtual environment. The outcome of this research is a tool that gives a virtual experience of typical situations of wheelchair users, and can also be used by people who have to design accessible environments.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 286–293, 2010.

2 A Virtual Wheelchair Simulator as a Serious Game

The application of simulation and virtual reality technologies can today be assessed as essential and cost-effective [5] [10] [11]. The term "Serious Games" specifies games that combine simulation and virtual reality with game play to train people in decision-making, problem-solving, and for other educational purposes in a "playful" way. The application of serious games has its origins mainly in the military [9] and the industrial sector. But also the medical and rehabilitation sector's attention is growing for the application of serious games. For instance, Jack et al. [4] are using virtual reality to help stroke patients in their recovery. The use of game engines as components off-the-shelf is a popular way to lower costs for developing virtual reality simulations or serious games [9].

The virtual wheelchair simulator described in this paper has been developed in cooperation with the Research Institute Technology and Disability [3] (FTB).During the development phase, an idea exchange took place between the two collaborating institutes, FTB and Digital Media Group at Bremen. Necessary data for implementation was either available at the FTB or could be "experienced" and researched in their laboratories. This included for example trying out different models of existing and commonly used electric wheelchairs. This close collaboration allowed the use of realistic data and the focus on requirements and needs of wheelchair users. FTB was furthermore capable of providing specification sheets of state-of-the-art electric wheelchairs that were essential for the implementation of a realistic driving behavior of the virtual wheelchair.

3 Driving Characteristics of a Wheelchair

The driving characteristics of an electric wheelchairare based on its respective purpose and application area. Wheelchairs that are predominantly used outside are preferably equipped with a direct steering technology and are designed to be driven at higher speeds without sliding in curve maneuvers. Wheelchairs primarily used for indoor driving purposes are usually equipped with an indirect steering technology and lower maximum speeds are required or even imposed to allow navigation in narrow and tight spaces. As a wheelchair driver usually needs to be able to use his wheelchair, both, inside and outside there are also hybrid models of wheelchairs that try to find a good compromise between the different requirements of indoor and outdoor wheelchairs.

Besides speed and maneuverability, wheelchairs have to fulfill safety characteristics like its minimum turning radius, an ability to get over barricades and obstacles, and tilt resistance in slopes and cross slopes. These characteristics are highly dependent on the construction and steering systems used in the different wheelchair models. For the research that we describe in this paper, the wheelchairs of the manufacturer Meyra have been used as a template. An overview about their specifications is given in table 1.

Table 1. Wheelchair models and their specifications engineered by Meyra

model	steering type	turn radius	max barricade	max slope	max cross slopes	tilt resistance
Clou	indirect	1 m max.	60 mm	12%	12%	15%
Compact 905	indirect	0,83 m max.	60 mm	15%	15%	15%
Allround 903	indirect	0,85 m max.	80 mm	15%	15%	15%
Allround 900 C	indirect	0,88 m max.	75 mm	15%	15%	15%
Optimus 2	direct	1,2 m	110 mm	18%	18%	28%
Touring 928	direct	1,2 m	110 mm	18%	18%	28%

4 Unit Conversion and Virtual Driving Characteristics

The simulation and transfer of a process from the physical world to the virtual world of UE3 requires to abstract all necessary data. The most basic unit that had to be conversed was a length unit for representing space in virtuality. Since the development of the first Unreal game and its release in 1998 the developers have been using the term "Unreal Unit" to refer to an elementary spatial unit. This specification has not changed since finishing the development of the game Unreal in 1998 and is still being used for determination of length etc. The application of Spring and Jennings virtualization taxonomy [7] led to the conversion of the length units as illustrated in table 2.

Table 2. Comparison: real world unit - virtual Unreal Unit

reality	virtuality
487.68 cm	256 UU
1 meter	52.5 UU
1 foot	16 UU
1 cm	0.525 UU
0.75 inches	1 UU

The current version of the wheelchair simulator allows steering the virtual wheelchair by keyboard and mouse or a gamepad. UE3 provides any kind of data input as the script engine includes classes that can handle input data from any interface. The use of any input device is possible that way, preferably a steering device should be chosen that is relatively similar to the "joystick"-like control devices used by real electric wheelchairs.

5 Assembling the Virtual Reality Wheelchair

UE3 allows a versatile use of self-assembled assets to augment the game or to create a game modification from scratch. The 3d model of the virtual wheelchair was created using Autodesk's 3d modeling software "3D Studio Max" and imported to the engine using a converter provided by Epic Games. To allow the wheelchair to be used in multiple scenario types, the base model of a hybrid usable wheelchair fits best to meet the requirements of diverse scenarios for a Serious Game. Therefore, the model of an indirectly steering wheelchair with rear-mounted drive has been chosen for conversion.

The selection of the camera perspective plays an important role for the player's immersion in the simulation. The used camera has its origin located according to the wheelchair driver's head position. This results in a first person view, which provides a stronger feeling of immersion to the player.

6 Scenario Development

The employment of the virtual wheelchair is predetermined for hybrid use. This allows to choose a scenario from an everyday situation in the life of a wheelchair driver: the access of a supermarket. Supermarkets should be accessible also for people with reduced mobility, e.g. users of rollators or wheelchairs. To guarantee a barrier-free access to such a public building the German Institute for Standardization (DIN) has approved several standards for accessing public buildings [2] and places [1]. The standardized values provided by the DIN have been abstracted to the engine by using the previously developed conversion units.

For constructing the virtual scenario, the level editor accompanying UE3, called "UnrealEd" was used. The supermarket has been created on an elevated stage to include the use of a wheelchair ramp. The scenario provides several challenges for a wheelchair driver. Narrow approaches like doors and close racks (cf. figure 1) demonstrate the difficulty that most wheelchair drivers have in everyday life, even accessing a place that is constructed with respect to established standards.

UE3 features numerous options in recording actions and events that occur in the engine. Several of these values were used to either adapt the physical behavior (driving characteristics) of the wheelchair or to detect important events like

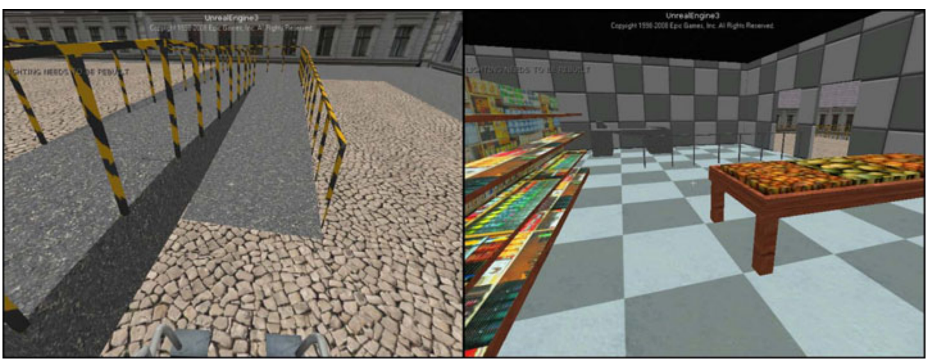


Fig. 1. Left: Accessing a wheelchair ramp with the VR-wheelchair; Right: Obstacles and narrow approaches inside the VR supermarket

collisions between the wheelchair and objects in the game world. The detected values can be used to trigger following events or simply be passed to the game console for monitoring. The use of recording and triggering events provides several options to add game play to the simulation and to make it a Serious Game in the end. A basic idea of the game play was the simple marking of the route that has to be accomplished by the virtual wheelchair driver in order to have a challenge. This route is marked with a row of red light spots (cf. figure 2), which turn green when they are approached correctly.

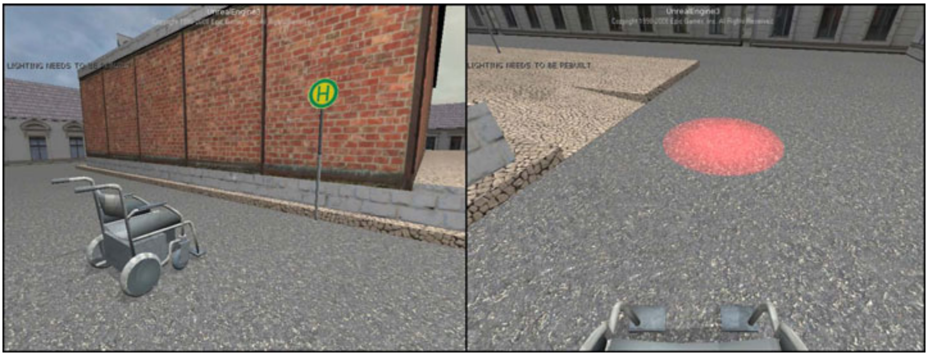


Fig. 2. Left: VR wheelchair in 3rd person view at starting point; Right: Trigger spot that turns green when correctly approached

7 Adaptation and Testing of Driving Characteristics

To achieve an as realistic as possible driving behavior of the virtual wheelchair, the driving characteristics and their appropriate conversion to the virtual domain were experimentally explored by systematically trying different parameter combinations and checking results for certain metrics and measurements related to the physical world, like turning radii. A highly experimental and empirical approach like this was required because no specifications of comparable values for the driving parameters of the virtual vehicle models of UE3 exists. For the adaptation of each driving characteristic from the physical world to the virtual world of the Unreal Engine a set of specific tests had to be developed. The guideline values for conversion were taken from real electric wheelchairs and are listed in table 1 in section 3.

7.1 Speed

Vehicles assembled in UE3 are provided with a parameter for speed. But as there is no factor yet for the conversion of a speed unit from a real world unit like kilometers per hour to the virtual world of UE3, a factor had to be determined experimentally. As time and length are the basis for this factor, a scenario was created, which was used to simply measure speed. For this, a game world was created in the UE3 level editor including two trigger points placed at the exact distance of 5250 Unreal Units, which is correspondent to a hundred meters in the physical world (cf. table 2). These two triggers write a time value to the console each time they detect a collision with the virtual wheelchair. The length of the time interval and the predetermined length of a 100 meters can now be used to calculate the conversion factor for speed from the physical to the virtual world by knowing the vehicles virtual speed parameter value. The measurements and calculations resulted in 14.50 Unreal Speed Units equal one kilometer per hour in the physical world. A speed of six kilometers per hour was determined to be applied to the virtual wheelchair because this is a typical maximum speed of electric wheelchair models.

7.2 Braking and Acceleration

Braking and acceleration are important factors to define a reliable handling and a wheelchair's safety. A vehicle in UE3 can be provided with several acceleration factors that are dependent on the current speed of the vehicle. This factor is used to adjust the acceleration behavior of the virtual wheelchair to that of its real equivalent. Smooth and controlled acceleration guarantees safe approaching to barricades for example. Even more important are braking maneuvers: the intention of the wheelchair driver to come to a full stop, i.e. maximum braking or “de-acceleration”, must not be that strong to toss the driver from the wheelchair seat, but the stopping distance must be as short as possible. The engine provides several torque and brake factors for vehicles that can be adjusted to behave like the braking of a real wheelchair.

7.3 Slopes and Cross Slopes

The safety approved values for a tilt of wheelchair models are between twelve and 18% pitch (cf. table 1). The absolute maximum average value for the tilting of a wheelchair is at 28%. The maximum value is higher due to a safety buffer before the wheelchair actually tilts over and endangers or even harms the driver. A realistic tilt value is desirable to make the simulation as realistic as possible. To adapt the tilting value to the virtual wheelchair model, the center of mass (COM) of the object was manipulated. UE3 allows one to move the center of mass in XYZ space. A movement of the COM in +Z or -Z direction which accords to “up” or “down” directions, leads to a change of the tilt value. To test the tilting behavior a row of ramps with increasingly raised pitches was created in the UE3 level editor. With these the tilting behavior could be adapted to the predefined values. Cross slopes are handled in the same way as values and parameters are the same except the wheelchair is standing transverse to the slope.

7.4 Cornering

A cornering can hold a dangerous situation for a wheelchair driver, especially when driving at higher speed rates. An electric wheelchair's steering behavior is usually linked to its speed. The safe turning radius dilates with the increase of speed. This behavior could be directly transfered to UE3 by having a parameter that allows the steering value to be adapted to the current vehicle speed.

7.5 Barriers and Maneuverability

Barriers and obstacles can be of a different manner. A curbstone for example can be surmounted with a wheelchair depending on its front wheel size. UE3 allows to adapt the height of all assembled wheels on the vehicle. This allows principally an adaptation of the model to any desired wheel height to make the virtual wheelchair surmount barriers that conform to standardized values. A cornering in a narrow corridor can also be considered as an obstacle. A wheelchair's turning radius significantly determines its maneuverability. The maneuverability of the virtual wheelchair has been tested by constructing circular platforms with different radii. Each wheel of the virtual wheelchair was monitored due to whether it is touching the ground or not. The test was invalid as soon as one wheel was off ground. A minimum turn radius of approximately two meters or 105 Unreal Units was ascertained for the virtual wheelchair. This big turn radius is due to the turning fulcrum lying on one of the rear wheels when turning. The overall width of the virtual wheelchair is 50.7 Unreal Units or 96 centimeters.

8 Evaluation of Driving Characteristics

The virtual wheelchair was assembled in UE3 by using a basis model that allows the creation of wheeled vehicles. The virtual wheelchair model was conceptualized as an indirectly steered wheelchair with rear mounted gear. Unfortunately, the indirect steering method could not be transfered directly to the engine. The physics engine did not allow indirect forces on the front wheels, which resulted in unexpected behavior in the simulation. A real indirect steering technology usually allows an individual torque control of each rear wheel and no torque on the front wheels. The different torque values are used for steering, thus the turning fulcrum is located in the middle between the rear wheels. This also affects the cornering radius significantly. With a correctly working physics model the cornering radius would be halved to approximately 52.5 Unreal Units or one physical meter (cf. section 7.5). The front wheels should also be able to turn freely through 360 degrees following the forces of the rear wheels. The conversion of a directly steered wheelchair model is the better choice when using a wheeled vehicle physics model from UE3 as the implemented model follows the behavior of a directly steered wheelchair.

9 Conclusion

The results of this research show how physics simulation in a game engine is used to create a virtual electric wheelchair simulation whose physical behavior is adequate to create realistic scenarios for everyday situations from the real life of wheelchair drivers. Engine specific tests were run to adapt the driving behavior to the one of a real wheelchair. As the used vehicle model of the engine did not fulfill all requirements to the chosen wheelchair model either another wheelchair model can be implemented or a different vehicle model in UE3 should be used in future work. Methods were applied that allow a correlated conversion of real units into the virtual system. We plan to employ our wheelchair simulator as a support tool in education of rehabilitation personnel and to use it as the basis for building serious games, probably in the form of an Unreal 3 "modification", around it to raise awareness for the difficulties of electric wheelchair users in everyday life.

References

1. DIN 18 024-1, 1998-01 Strassen, Plätze, Wege, öffentliche Verkehrs- und Grünanlagen sowie Spielplätze (1998), http://nullbarriere.de/din18024-1.htm (accessed on March 10, 2010)
2. DIN 18 024-2, 1996-11 Öffentlich zugängige Gebäude und Arbeitsstätten, http://nullbarriere.de/din18024-2.htm (accessed on March 10, 2010)
3. Forschungsinstitut Technologie und Behinderung. Website, http://www.ftb-net.com (accessed on March 10, 2010)
4. Jack, D., Boian, R., Merians, A., Tremaine, M., Burdea, G., Adamovich, S., Recce, M., Poizner, H.: Virtual reality-enhanced stroke rehabilitation. IEEE Transactions on Neural Systems and Rehabilitation Engineering 9(3), 308–318 (2001)
5. Niniss, H., Nadif, A.: Simulation of the behaviour of a powered wheelchair using virtual reality. In: 3rd International Conference on Disability, Virtual Reality and Associated Technologies 2000, pp. 9–14 (2000) (citeseer)
6. Sawyer, B., Smith, P.: Serious games taxonomy. In: Slides from the Serious Games Summit at the Game Developers Conference (2008)
7. Spring, M., Jennings, M.: Virtual reality and abstract data: Virtualizing information. Virtual Reality World 1(1) (1993)
8. Stredney, D., Carlson, W., Swan II, J., Blostein, B.: The determination of environmental accessibility and ADA compliance through virtual wheelchair simulation. Presence-Teleoperators and Virtual Environments 4(3), 297–305 (1995)
9. Susi, T., Johannesson, M., Backlund, P.: Serious games–An overview. Skövde: University of Skövde (Technical Report HS-IKI-TR-07-001) (2007)
10. Virvou, M., Manos, K., Katsionis, G.: Virtual Reality edutainment: cost-effective development of personalised software applications. In: Gavrilova, M.L., Gervasi, O., Kumar, V., Tan, C.J.K., Taniar, D., Laganá, A., Mun, Y., Choo, H. (eds.) ICCSA 2006. LNCS, vol. 3980, p. 251. Springer, Heidelberg (2006)
11. Voas, J.: Cots software: the economical choice? IEEE Software 15(2), 16–19 (1998)

Event-Based Data Collection Engine for Serious Games

Amith Tudur Raghavendra

Carnegie Mellon University
Entertainment Technology Center
700 Technology Drive Pittsburgh, PA-15219 310-910-4274
amith@cmu.edu

Abstract. Games with a purpose other than entertainment can be called Serious Games. In this paper, we describe a generic event-based Data Collection Engine (DCE) that has been developed for Serious Games on the Unity Game Engine. Further, we describe a framework that allows for the manipulation and feedback of the collected data back into the game in real-time. The player experiences the visuals, sounds and the game itself that is streamed over the web. The player engages with an enriching, multimedia experience allowing him/her to be immersed in the game. By suitably designing the serious game we could determine the behavior of the player in real world under the given scenario or other scenarios. The DCE is optimized to collect relevant data streamed online without affecting the performance of the game. Also, the DCE is highly flexible and can be setup to collect data for any game developed on the Unity Engine.

Keywords: Unity Game Engine, Serious Games, Data Collection Engine.

1 Introduction

We have developed a system that can collect data from Serious Games[1], while they are being played. The event-based data collection is efficient and can be integrated into any game that has been developed on the Unity Game Engine [2]. Data collection from a Serious Game played online can quickly become immense and voluminous. It would need an efficient approach that would stream both the game as well as the collected data. It would have to be flexible enough to include the different kinds of games that can be created and different data that might have to be collected on a single game. In the sections below we describe a generic event-based data collection approach that can be used to track and analyze data in any Serious Game developed on the Unity Game Engine. The system is designed to track only the data that is relevant, optimizing the data flow for the game played on a browser. Additionally the DCE has the capability to manipulate the collected data and feed it back to the game in real-time. So a game can be customized dynamically to a player based on the previously collected data. The DCE could be used to single out and profile a player over multiple games.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 294–301, 2010.

2 Scenarios

Scenarios in Unity are made up of game objects. The game objects could consist of the multimedia like sound spots, animations, models and characters that are interactive. Figure 1 shows one such scenario. The game objects are circled red. Each of these game objects are made up of parameters that determine their state at any instant in the game. The intuition behind the DCE would be the fact that by logging and correlating the states of the parameters at a particular instant we could analyze the scenario in several different ways. Let us take the example of a Market Scenario. We have vendors along the sides of a road. The player has to drive a car through the Market streets. Here the game objects would include the vendor, the items they are selling, road, pavement, car and the driver (player). The player's parameters would include position, view of the market, sound being heard, etc.

Fig. 1. A Scenario in the Unity Game Engine

3 Generic Data Collection Concept

Generic Data Collection is based on the concepts of Condition, Selection and Events. As mentioned earlier it relies on the intuition that by tagging the data suitably, the parameter values can be correlated later and the scenario as a whole could be analyzed to any depth of complexity.

3.1 Condition

Time is abstracted into the concept called a *Condition*. A *Condition* determines *when* a parameter(s) would be logged. The *Condition* can be a relation between multiple objects that may turn out to be true in one or more frames in the game. For example, in the market scenario if we want to know about the case of a reckless driver, the condition would be track *when* position of the car is on the pavement and the velocity of the car is greater than 40mph.

3.2 Selection

The parameters to be logged are abstracted into the concept called a *Selection*. A *Selection* determines *what* would be logged. For example, in the market scenario, suppose we wanted to log the positions of the vendors when the car runs over the pavement and the velocity/path of the car. Then, the parameters to log would be the positions of the vendors, position of the car and the velocity of the car.

3.3 Event

By combining a *Condition* and *Selection* together we get an *Event*. *Events* greatly reduce the complexity of data collection in a world composed of hundreds of objects continually interacting with each other. We check for an *Event* in every frame of the game. An *Event* occurs when a *Condition* is found to be true, in which case we log the *Selection*.

Figure 2 shows the generic concept diagram for the Event handling. *LogEventChecker* looks for an event every frame. This could be customized to any granularity based on the

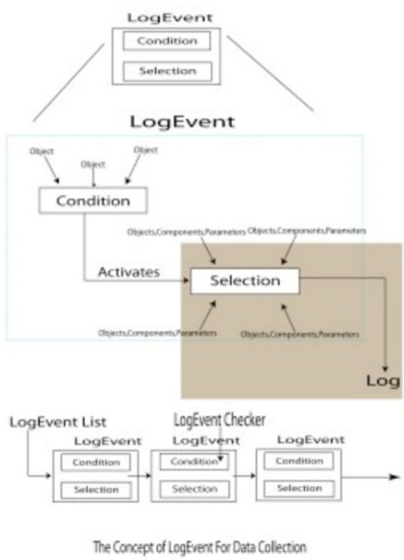


Fig. 2. Concept of Generic Data Collection

capability of the hardware at hand. However if the data is collected, say every 24 frames, then the trade-off would be the precision of data analysis later on. Once the *LogEventChecker* finds that a given *condition* is satisfied, it triggers a *selection* to be logged. The parameters are collected in a set to avoid redundancy. Finally, the set of all parameters for this frame are logged onto the database.

For example, in the market scenario, an event is composed of the condition and selection that were mentioned above. Hence, when the car runs over the pavement, the condition is met and hence a log is created with the parameters in the selection. The parameters are tagged by a unique gameid and a frame number. Hence the primary key in the database would be (gameid, frame number). The correlation would involve querying the values of the parameters at a particular frame number. In this case we could query for the values of the parameters when the car went over the pavement. This gives us an idea about the game when the event under consideration happened.

3.4 TARDA

Test Application for Reactive Data Analysis (TARDA) is a module attached to the database that feeds the data back into the game changing it in real time. The user could write the relevant analysis logic [3] in the TARDA by utilizing the huge amount of data that is present in the database as a result of the players' previous games. For example, let us take the market scenario again. The Game ID would be a unique ID defining the particular game instance for a particular user and can be used to track a person over multiple sessions of the game. If a particular driver is noted to be reckless over a bunch of games, let's say 20, then we could modify the behavior of the vendors in the 21st game. We could program them to walk away from the car as it approaches. As soon as the car moves over the pavement, the vendors can take the path that the driver is least likely to take. The least likeliness could be determined by using the path data of the car that has been collected from the past 20 games. Thus the game can be programmed to change dynamically by getting its feed from TARDA. The intelligence of the person playing the game can be propagated to the system through TARDA over time. We could mine the data across multiple users over thousands of games that are played by them. For example, we could figure out the expected path of car from the past 2000 games from the most recent 100s of users.

4 The Data Collection Framework

Figure 3 shows the Data Collection framework which includes the implementations for the *EventChecker* and TARDA. We create the game on the Server establishing the different parameters that would have to be collected at different instants. A customized and streamlined GUI has also been developed to facilitate the Event Creation. Multiple Events that are created in the process are stored on an *EventFile* on the Server. The build is then published and deployed on a web browser. A player could play the game from any part of the world since the game just requires a Unity plug-in. The player can be uniquely tagged over multiple sessions of the game, as data about him/her is gathered by the DCE. All the while, the player is oblivious to the fact that the DCE is working behind the scenes. TARDA that is attached to the Database can be used to tweak the game in *real-time* based on the user's profile that has been gathered over a period of time and this game as well. The Database has multiple tables that are used to store the data. The DATALOG table stores the (key, value) pair that is being logged. TORUN is the database that keeps track of the TARDA commands. These commands can be fed into the game while its running. The game will change its state depending on the command that has been fed in.

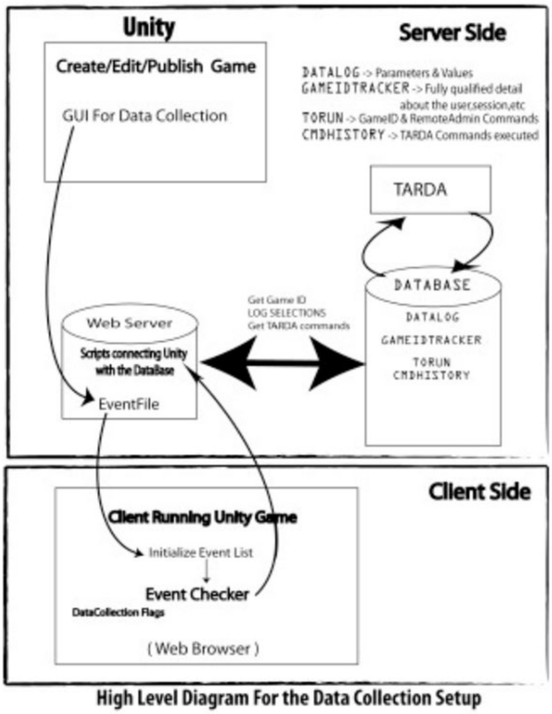


Fig. 3. The Data Collection Framework

5 Implementation Examples

5.1 Editor GUI

Figure 4 shows the game developer building the events with the customized GUI that has been created. The creator goes through the logical process of creating a condition and selection here. Condition is a code in the native language of the game being created for which there will be no learning curve. As it can be seen here, we have selected a person. The person can be referenced using the code that is written in the GUI. Similarly we could have a bunch of objects to be logged when this condition is met. The process is repeated for every event that the developer intends to track. The code written here is evaluated dynamically at runtime. Hence this also provides a mechanism for TARDA to introduce new methods at runtime into the game.

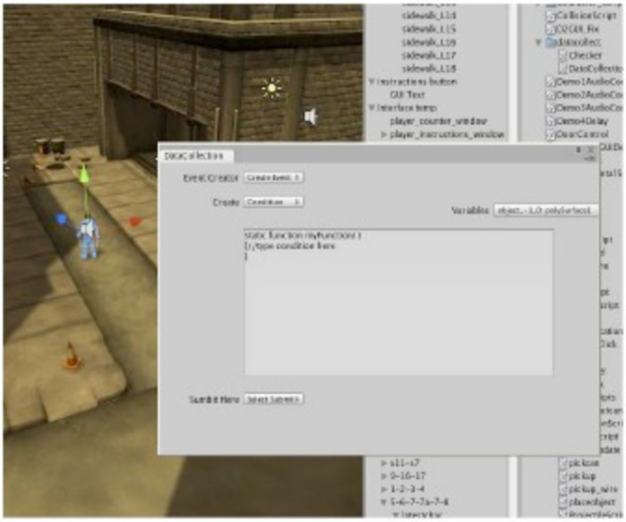


Fig. 4. Create the Events

5.2 Position Tracking

Figure 5 shows a simple example of a game involving a person running in a virtual world collecting ammunitions. This is the top view (map) of the world. The DCE was added to it seamlessly and the person's positions as well as the places in map where he collected the ammunitions were collected into the database.

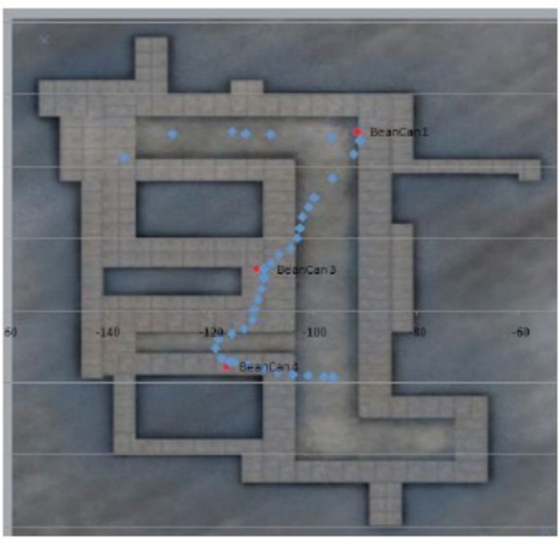


Fig. 5. Position Tracking from the Database

5.3 TARDA Feedback

Figure 6 gives an example of TARDA where we have the person's position being fed back from the database. Game ID of the current game is being shown as well. TARDA has access to the entire database; hence the feeds could be from any of the game that this player or any other player has experienced.

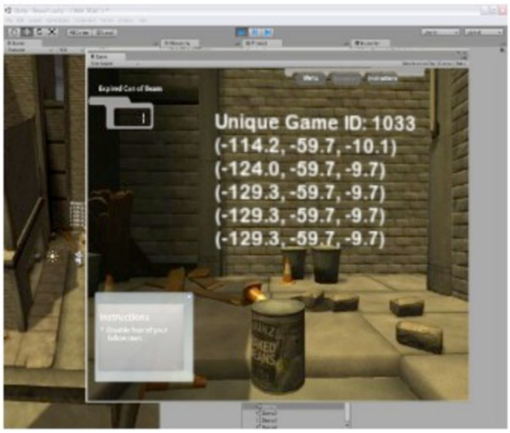


Fig. 6. TARDA feedback into game

Figure 7 gives an example of another TARDA feedback. Here the lines and arrows are pulled into the current game from the database. These lines represent the path taken by this player 10 games before. TARDA feedback is dynamic and lines from games being played by different users can be manipulated independently and concurrently fed into this game.

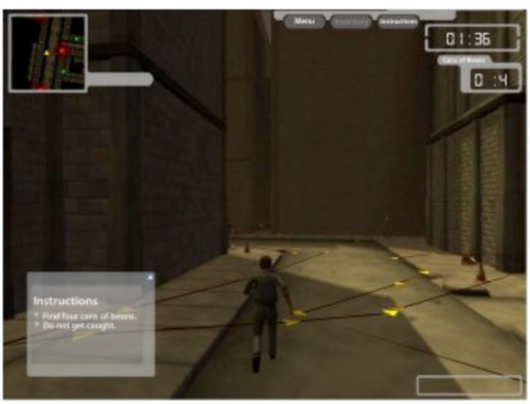


Fig. 7. TARDA feedback into game

5.4 Dynamic Tracking

The lines in Figure 8 show the position of a person/car over time. The yellow line is being drawn dynamically as the game proceeds. The black lines represent the path from the previous games. This is a simplified view of the data that can be gathered from the game.

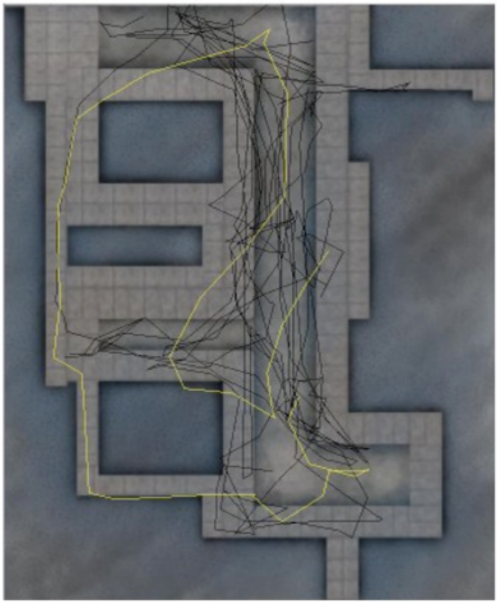


Fig. 8. Position of the player over time

Acknowledgments

My thanks to John Conomikes and Matthew Crozier for being part of the work. Drew Davidson, Mike Christel and Scott Stevens for the guidance. Thanks to Don Marinelli, Gary Sarkesian, Adam Pierce, Shirley Saldamarco and James Burke Jr. for the support. Most importantly, thanks to Austin Jephson, Srinavin Nair, Ethan Rupp, Akash Joshi and Hee Jun Kim for being part of our development team.

References

1. Michael, D., Chen, S.: Serious Games. Thomson Publishers (2006) ISBN:1-59200-622-1
2. Unity Game Engine, `http://unity3d.com/`
3. von Ahn, L., Dabbish, L.: Designing games with a purpose. Communications of the ACM 51(8) (2008), ISSN:0001-0782, `http://doi.acm.org/10.1145/1378704.1378719`

Culturally Sensitive Computer Support for Creative Co-authorship of a Sex Education Game

Junia C. Anacleto[1], Johana M.R. Villena[1], Marcos A.R. Silva[1], and Sidney Fels[2]

[1] Advanced Interaction Laboratory, Department of Computer Science
Federal University of São Carlos, Brazil
{junia,johana_villena,marcos_silva}@dc.ufscar.br
[2] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada
ssfels@ece.ubc.ca

Abstract. We describe a computer-supported game authoring system for educators to co-author a game to help teaching sensitive content, specifically sex education. Our approach provides educators the ability to co-author the game to tailor it to the class based on a computer-supported interface that draws upon a large, cultural database. By targeting the game to the culture of the students, they feel their values, beliefs and vocabulary are being considered in the game, providing better comprehension of the content, leading to stronger learning engagement that is helpful for highly charged, sometimes uncomfortable and sensitive material such as sex. We studied our design in the classroom and observed that giving educators co-authorship of the game helps them adopt using online games.

Keywords: Cultural Context, Common Sense, Game, Education.

1 Introduction

Sex education often is a difficult subject to teach due to factors such as the teacher can be embarrassed about the material, the students are embarrassed, the vocabulary, generation and cultural gaps between teacher and student interfering with understanding and learning, and ensuring that the material covered is appropriate to the class dynamics and background. Games have the potential to help with this material as they can provide a fun, non-confrontational, low emotional investment environment to cover sensitive topics [5]. However, traditional games are static, thus may not be appropriate for a particular class and its cultural context, especially for topics that fluctuate socially such as attitudes and knowledge about sex. The challenge for tailoring game content though is that considerable creativity and knowledge of the people who will be engaged in the game is required. Our computer-support draws upon a very large database of cultural knowledge that contains information about the attitudes, vocabulary and knowledge associated with different demographics allowing the system to supply ideas to the educator to work into the game. Our contention is that the targeted support of educator's creativity within a well-defined type of game leads to a rich, educational experience for the students, connecting them closer to the educator.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 302–307, 2010.

The main contributions this paper describes include: (1) a co-authored online sex education game, called *What is it?*, that can be used effectively; (2) a generalizable approach to integrating cultural knowledge into a computer supported game authoring interface to allow educators to be creative to adapt games to their students; and (3) evidence that educators more readily adopt online games in the classroom if they have control over the content through co-authoring it. We call the tailoring process, co-authoring, as the main game structure has been created by someone other than the teacher, however, they are adding content to the game which suggests that the final game has been co-authored. We discuss the related work in Section 2 that focuses on cultural support for games. Section 3 describes the sex education game. Section 4 discusses the evaluation of this game in the classroom. We then conclude the paper and provide future work directions in Section 6.

2 What Is It?

We are using common sense statements as cultural knowledge to give cultural context to the sex education game the educators develop. Common sense is the knowledge shared by most people in a particular culture [1]. As complexity of computer applications grows, one way to make them more helpful and capable of avoiding unwise mistakes and unnecessary misunderstandings is to make use of common sense knowledge in their development [5]. For our sex education game, we want to extract the knowledge appropriate for the students who will be playing the game so that the teacher can use it effectively.

We use the OMCS-Br project (Open Mind Common Sense – Brazil) to collect common sense of the Brazilian general public. Within the OMCS-BR database are encodings of the relationships between statements generated by people of all ages about attitudes and beliefs around sex. This database provides the relationships about sex related terms that are useful for providing words and concepts to the teacher to be creative when she is tailoring the sex education game [1][3]. Using it, she can ensure the content reflects the cultural reality of the students' sexual knowledge in a fun game.

We illustrate the potential of our approach by adding cultural support to a sex education quiz game called, *What is it?* The student's interface is shown in Figure 1. *What is it?* is a web quiz game where the player sees a topic with an associated secret word. Their task is to guess the word based on a series of clues that are shown one at a time. The objective is to guess the word after seeing the least number of clues. There are ten clues that can be selected by the learners by clicking on a number. The challenge for the teacher is to make a set of secret words and clues that are related to the students' concepts and vocabulary, otherwise, they will be alienated from

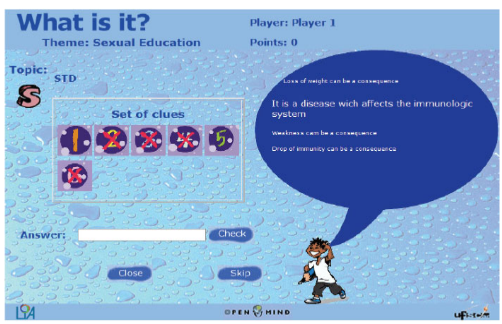


Fig. 1. Student interface to the *What Is it?* sex education game

the concepts being described. For example, many sex education curricula are taught in biology classes using scientific names for everything to sterilize the content. However, this does not allow students to assimilate the knowledge in a relevant way to their current situation as they don't make the connection between the impersonal scientific vocabulary and their reality.

We chose sex education as our target since in Brazil, teachers must cover the transversal themes defined by Fundamental Education Secretary/Minister of Education (SEF/MEC). These themes aim to promote citizenship upbringing in school. The themes are: Ethics, Health Care, Environment, Sexual education, Cultural diversity, Work and consumers [4]. However, teachers often find they are not provided adequate training to teach the sex educational components and often are very uncomfortable covering the material. Thus, the main objective of this game is to make possible the use of cultural knowledge in web educational games, stimulating the introduction, reinforcement and the knowledge's construction in the learning process to allow teachers to more effectively cover these areas.

2.1 Culturally Sensitive Authoring of the Sex Education Game

We provided a computer-supported teacher interface to allow the teacher to define the set of secret words and the clues that would be associated with them. There are 10 clues needed for each secret word which becomes quite a creative challenging to come up with good clues and secret words. It is here where the OMCS-Br cultural knowledge can help support this creative process. The clues for the game are defined using the support of the OMCS-Br cultural knowledge base that helps the teacher adopt a common vocabulary to the students and clarify myths, beliefs and taboos culturally inherited.

There are seven steps to support a teacher to create a quiz: (**1**) the teacher first selects the profile of the learner, age range and region of the country to identify what group of learners the configuration is more appropriate; (**2**) the teacher defines the game's main theme, such as sex education or health, that is used to constrain the cultural knowledge base search; (**3**) some of the related topics to the main theme are shown in the teacher interface (shown in Figure 2) so that the teacher can decide which topics he wants to explore in the game; **(4)** a list of possible secret words is shown based on the search results allowing the teacher to choose one or create one of his own; (**5**) a list similar words is shown based on the secret word in the previous step. The teacher can also insert synonyms for secret word selected; (**6**) the teacher defines 10 clues; clue suggestions come from the cultural knowledge base that he can select from or he can edit it, or ignore it and create his own. The suggestions are intended to trigger creative ideas for the clues as well as provide some choices that might be useful. The selected clues are shown in region II of the interface shown in Figure 2; and (**7**) Once complete, the teacher sees all clues and the secret word for final editing and confirmation.

In a separate view, the teacher highlights the relevance of the cultural knowledge as a factor that interferes in student's learning process. This approach brings discussions about what was previously described, like myths, beliefs, misunderstandings into the classroom so these ideas won't be understood as facts by the students.

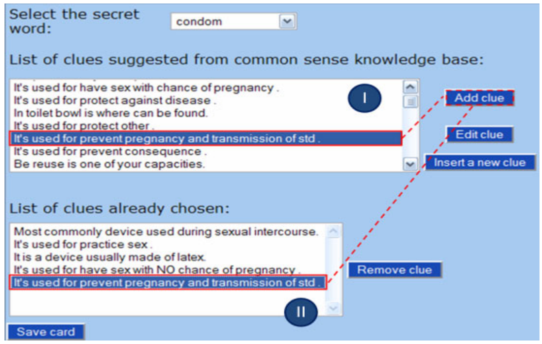


Fig. 2. Clues suggested from cultural knowledge base for the 'What is it?' game

The teacher repeats all seven steps for each secret word that they want to include in the game. Once complete, the students can then go to the website to begin playing the game to guess the secret words from the clues. Notice that the teacher can include misconceptions that student may have that have appeared in the database as these typically are captured. For example, if the secret word is HIV/AIDS, one clue could be, "You can't get it from a public toilet." as the contrary may have been in the database. Interactions with the game are logged so that the teacher can see where students are having trouble for further focus on those concepts. In the end, teachers can consider and use the cultural knowledge as needed from the computer support in their pedagogical activities aiming at helping the learners to associate the content being taught to their personal experiences and knowledge [1].

We studied this web-based sex education game in a school as described in the next section. Of particular interest are two main design issues: (1) how teachers responded to having a role to play in co-authoring the game; and (2) whether the cultural knowledge used in the computer supported co-authoring stimulated their creativity and/or was useful in some way.

3 Evaluation

We used a case study approach to evaluate this game because we wanted to evaluate the game in a real-life context, in this case, the classroom. According to Yin [6], case study is an effective type of empirical research that investigates a phenomenon embedded in a context of real life. For each classroom we studied, we used the following data collection techniques: questionnaires pre-section and post-section based on QUIS [2], video recording and screen capture and direct observation with four researchers. We evaluated *What is it?* in two classrooms. The researchers selected six teachers and nine students. In the first classroom, there were four students in the age range of 11 years old and they played the sexual education theme. In the second classroom, there were five students in the age range of 12 until 15 years old and they played the health care theme. Teachers were asked to create a new instance of a game, choosing any of the themes and topics. They used the system for approximately 40 minutes to set up a new instance of the game with twenty clues and two secret word; the themes used in these settings were: environment (3 times), healthcare (twice) and sexual education (once). After this step, the students were asked to play the instances. Each player used the game for about 50 minutes, playing an average of 7 times different instances of the same theme.

3.1 Results and Discussion

After the teachers created the instance of the game, they received a post-session questionnaire. Each point represents an alternative which shows the teacher's opinion. For example, in the first question, "What did you think about the game creation process?" the alternatives were: I have no opinion, Very hard, Hard, Indifferent, Easy, Very easy from zero to six. For this particular question, just one teacher said that this support was unhelpful or irrelevant, in which she commented, "I thought it was very unproductive to have to read a huge list of clues with many phrases" and "It is hard to find a clue that fits well to the secret word". However, for the others, the game process of co-authoring was simple; all these teachers had a good experience to create their own game. Table 1 shows three post-session questions with the teachers' answers.

Table 1. Teachers´ Post-session questionnaire answer about the creating of the game. (Translated from Portuguese, 5 is strongly liked.)

Questions	Teacher 1	Teacher 2	Teacher 3	Teacher 4	Teacher 5
What did you think about the creating game process?	easy	hard	easy	easy	easy
What did you think about the game, created by you, to work the themes with students?	very inadequate	appropriate	appropriate	very appropriate	appropriate
What did you think about the quiz game strategy, i.e., guess secret word taking into consideration the clues, created by you?	very inadequate	indifferent	appropriate	very appropriate	appropriate

We also asked the students if the language of the game was familiar to them, which they agreed with. Of course this does not mean that it was only due to the common sense knowledge support, however, as this support was used by the teachers, we believe that our approach helps to achieve this goal. We noticed that the students did understand that material as well. For example, there was a question "Did you get to understand the clues?" for which one student said "always", five said "sometimes", one said "indifferent", two said "rarely".

The results confirm that the "What is it?" environment provides extra benefits beyond a simple educational game. The teacher's ability to be the co-author of the game content is considered beneficial by most of the teachers allowing the game to be tailored to different students with different issues. In order to support the teachers during this process, the software uses the common sense knowledge base to provide some suggestions of clues for each card, which were generally accepted by the teachers. Almost all teachers used part of these suggestions in order to create their own clues.

Our experiment shows that, except for one teacher, the common sense knowledge base helped with creating the game content. The criticism made by the one teacher may be interpreted in two different ways: (1) since the common sense knowledge base content is created by any person who accesses the project site, there are many concepts that do not make sense or that are scientifically wrong and do not provide meaningful options when presented as a choice. It is the teachers' decision how they could work with this information; (2) the process to create the clue has some problems and can be improved in order to create better sentences. We are addressing the second interpretation by working with a natural language processing research group to improve the sentences suggested to the teacher. Nonetheless, the suggestions allow the teachers to successfully create clues for the secret words.

4 Conclusions and Future Works

We created a new web-based sex education game to illustrate that cultural knowledge can be included in computer games to support the creative process of co-authoring the game in and educational context. Our approach is general and can be applied to many games where the interface needs to be authored to take into account the cultural context of the players. Through our deployment our main observations suggest the following: (1) the ability to co-author the game through the tailoring system we built empowered the teachers and was very successful. Teachers felt that this made the game useful in their setting and would use it again because of that; and (2) the cultural knowledge component was found to be useful, but it is still premature to determine whether it will be accepted as part of the methodology for creating culturally sensitive games.

Thus, the ability to co-author the games was a helpful element that made the game successful. The fact that teachers indicated they would like to continue using it because of that and that they would explore the potential of the computer-support system more as they created more games suggests that there will be an important role for our approach. We are continuing to explore this promising direction.

References

[1] Anacleto, J.C., Lieberman, H., Tsutsumi, M., Neris, V.P.A., Carvalho, A.F.P., Espinosa, J., Zem-Mascarenhas, S.: Can common sense uncover cultural differences in computer applications? In: Bramer, M. (org.) Artificial intelligence in theory and practice - WCC 2006, vol. 217, pp. 1–10. S-V, Berlin (2006)

[2] Chin, J.P., Diehl, V.A., Norman, K.L.: Development of an instrument measuring user satisfaction of the human-computer interface. In: Proceedings CHI 1998 (1998)

[3] Ferreira, A.M., Pereira, E.N., Carelli, I.M., Anacleto, J.C., Silva, M.A.R., Dias, A.L.: A culturally Contextualized Web based Game Environment to Support Meaningful Learning. In: International Conference on Computer Supported Education (CSEDU 2009), Lisboa (2009)

[4] Liu, H., Singh, P.: ConceptNet: a practical commonsense reasoning toolkit. BT Technology Journal 22(4), 211–226 (2004)

[5] Resnick, H.: Human services Technology Innovations in practice and education. Harworth Press (2002)

[6] Yin, R.K.: Case Study Research. Design and Methods. Applied social research method series. Sage Publications, California (2002)

Real-Time Caustics in Dynamic Scenes with Multiple Directional Lights

Budianto Tandianus, Henry Johan, and Hock Soon Seah

School of Computer Engineering, Nanyang Technological University,
N4 Nanyang Avenue, Singapore 639798
{budi0010,henryjohan,ashsseah}@ntu.edu.sg

Abstract. We present a real-time GPU caustics rendering technique in dynamic scenes under multiple directional lights taking into account light occlusion. Our technique renders caustics cast on receiver objects as well as volumetric caustics. We precompute caustic patterns of caustic objects for several directional lights and store them in caustic images. During the rendering, we interpolate the precomputed caustic patterns based on a given light direction. One of the applications of our technique is to render approximate caustics under environment illumination. To achieve this, we propose an environment cube map segmentation technique which divides cube maps into several light regions with each region is represented using one directional light.

Keywords: Caustics, Real-Time Rendering, Environment Illumination, GPU.

1 Introduction

Real-time photo-realistic rendering is a major goal in computer graphics and entertainment as it can help the audience to experience or even immerse into the virtual world as if it is the real world. However, it is computationally expensive to generate photo-realistic images. One of the important effects in photo-realistic rendering is caustics which are produced by reflective and/or refractive objects (**caustic objects**).

In this paper, we present a real-time caustics and volumetric caustics rendering technique under multiple directional lights taking into account light occlusion. Figure 1 shows some examples of our caustics rendering results. Our rendering technique can also be used to render approximate caustics under environment illumination. We achieve this by approximating the environment illumination using a set of directional lights computed using our proposed environment cube map segmentation technique. We precompute the reflective and/or the refractive caustic patterns at the surrounding of caustic objects based on a set of directional lights. We use the precomputed caustic patterns in the rendering pass in order to efficiently compute the caustic intensities at arbitrary locations. Our proposed technique has some differences with the technique presented by Wyman et al. [8] and we discuss these in Sections 2 and 3.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 308–316, 2010.

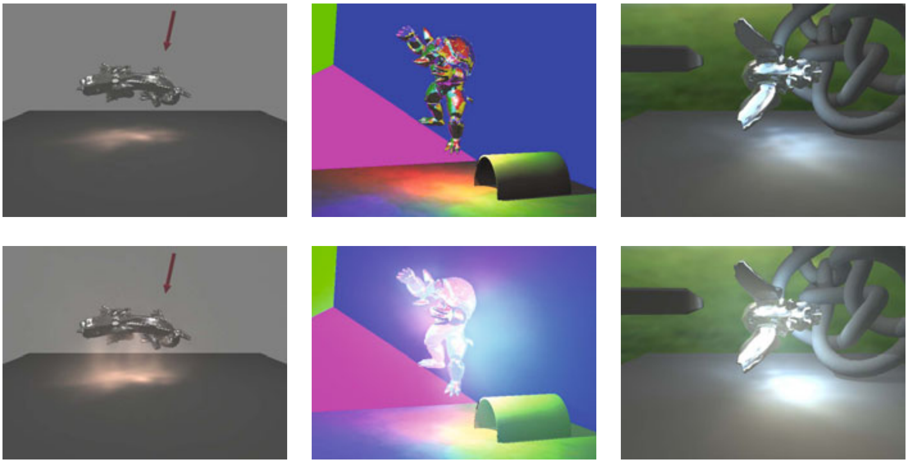


Fig. 1. Caustics rendering using our proposed technique. First column shows caustics under one directional light source (light direction indicated by the arrow). Second and third columns show caustics under environment illumination. The first row shows only the cast caustics, and the second row shows cast caustics and volumetric caustics.

2 Related Work

In photon mapping [4], photons (packets of light energy) are shot to the scene and stored in a photon map. The photon map is used during rendering by gathering the photons around a visible point in order to estimate the caustic intensity. However, it is computationally expensive and it needs to be recomputed if there are any changes in the scene. To accelerate the photon mapping, Günther et al. [2] use a computer cluster and they are able to achieve interactive rates whilst Purcell et al. [5] use a GPU and they are able to reduce the rendering time into few seconds. Zhou et al. [11] propose a real-time GPU-based kd-tree generation technique which greatly aids the photon mapping process. However, the overall rendering speed is mostly below 10 frames per second.

Another approach is the image-based caustics rendering proposed by Shah and Koninen [6], Sun et al. [7], Wyman and Davis [10]. In this approach, they shoot photons through each pixel by rendering the scene from the light source. In many cases, these techniques only support one light source to achieve interactive or real-time performance, do not support total internal reflection, and cannot generate volumetric caustics.

Wyman et al. [8] precompute the local caustics of a caustic object on uniform grids or concentric spheres (with constant radii differences between the consecutive spheres) enclosing the caustic object. They use a CPU cluster for both precomputation and rendering (their rendering speed is below 20 frames per second). In the rendering, the scene is illuminated by point or directional light sources, not environment illumination. On the other hand, our technique renders real-time caustics under environment illumination using a GPU.

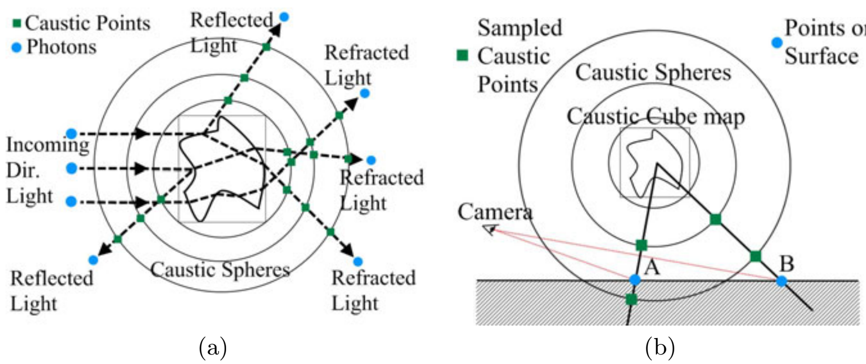


Fig. 2. (a) shows the precomputation of caustic patterns. The incoming directional light (implemented as photons) is reflected and refracted by the caustic object. (b) shows how we compute the caustic intensities on the visible points by interpolating or extrapolating the sampled points of nearby caustic spheres.

3 Precomputing Caustic Patterns

The purpose of the precomputation is to record the caustic intensities at the surrounding of caustic objects. We precompute the reflective and refractive caustics of a caustic object using photon mapping [4] on a set of concentric spheres (**caustic spheres**) based on a set of directional lights as illustrated in Figure 2(a). Unlike Wyman et al. [8] who linearly change the radii of the spheres, we use a quadratic function to determine the radii of the caustic spheres since caustics weaken quadratically because of light attenuation. By using a quadratic function, we densely sample near to the caustic object and sparsely sample far from the caustic object. As a result, we can reduce the number of caustic spheres while maintaining the visual quality. Specifically, we use the following equation to compute the radii.

$$r_i = r_{\min} + (r_{\max} - r_{\min})((i-1)/(s-1))^2, \tag{1}$$

where r_i is the radius of the i-th caustic sphere, $r_{\min}$ is the minimum radius, $r_{\max}$ is the maximum radius, and s is the number of caustic spheres. We set $r_{\min}$ to be slightly greater than the distance from the center of the caustic object to the nearest surface of the caustic object. Assuming the maximum dimension of the object's bounding box d is max{width, height, depth}, we set $r_{\max}$ to $4d$ based on our experiments in which caustic patterns were barely noticeable on the caustic spheres with the radii of $4d$ and more.

4 Rendering Caustics under Directional Lights

Single Directional Light. Given an arbitrary light direction L, we choose the four precomputed light directions nearest to L, rotate the caustic spheres of those light directions to align their directions with L (similar to Wyman et al. [8] who use three nearest light directions) and blend them using bilinear interpolation. Afterward, we use the computed caustic spheres of L to determine the caustic intensity at points in the scene. The caustic intensity at a point is trilinearly interpolated (point A in Figure 2(b)) or extrapolated (point B in Figure 2(b)) using the intensities at the eight nearest samples in the caustic spheres of L.

Caustics cast on receiver objects. To render these caustics, we first compute a **caustic cube map** (Figure 2(b)) which stores the information of receiver points (the coordinates and depth value of the points) as well as their caustic intensities. Similar to shadow mapping, we compute those values by rendering the surrounding scene from the center of the caustic object and the caustics in the caustic cube map are then projected to the scene.

Volumetric caustics. In the presence of participating media (assumed to be homogeneous), we generate volumetric caustics by casting a ray from every pixel on the screen to the scene [3] and we integrate the caustic intensity at each sample point on the ray.

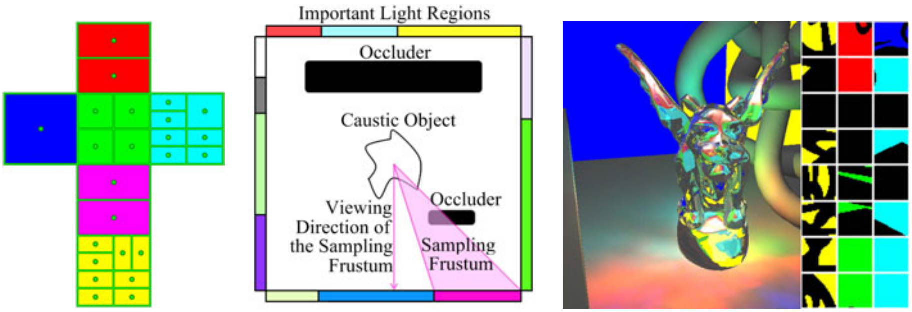
Multiple Directional Lights. To generate caustics under D numbers of directional lights, we apply the algorithm for one directional light to each light and accumulate the caustic patterns on the caustic cube map.

5 Rendering Approximate Caustics under Environment Illumination

One of the applications of our caustics rendering technique is to approximate caustics under environment illumination. In this case, we represent the environment illumination as an environment cube map. Since integrating the radiance from all pixels in the cube map during rendering is impractical, we segment the environment cube map into several important light regions and represent each of them with a directional light. Thus, rendering caustics under environment illumination is similar to the rendering under multiple directional lights (Section 4), that is integrating caustic patterns from all directional lights.

Environment Map Segmentation. Debevec [1] recursively segments the environment map (latitude-longitude format) into two regions having almost equal total radiance until a number of iterations. In our technique, to account for light occlusion, we need to represent the environment illumination as a cube map. Therefore, we cannot directly use Debevec's segmentation algorithm [1].

In our proposed environment cube map segmentation algorithm, in each segmentation iteration, we choose the region with the most radiance and segment it into two new regions having the same total radiance (similar idea as Debevec's segmentation [1]) until we obtain D light regions. We start the segmentation from six regions corresponding to each face of the cube map. For each light region, we store the weighted center (with the weight is the radiance of each pixel) of the



(a) (b) (c)

Fig. 3. (a) The result of our environment cube map segmentation with the lines are the region boundaries and the dots are the light directions representing the regions. (b) shows how we sample the important light regions for their radiance contribution by rendering them taking into account occluders (rendered as black color). (c) shows a scene configuration using the environment map in (a) with the sampling results shown on the right side. Note the occluders are rendered as black objects.

region as the important light direction along with the region boundary information. Figure 3(a) shows the results of our environment cube map segmentation algorithm.

Directional Light Radiance Sampling. As the directional lights are derived from an environment cube map, the radiance of each directional light is the total radiance of all unoccluded pixels in the important light region (corresponding to the directional light). To sample the regions by rendering, we set up the sampling frustum by using the boundary information of the important light regions (obtained from our environment cube map segmentation algorithm). As we sample the regions by rendering, we render the surrounding objects as black color in order to take into account light occlusion. Figure 3(b) illustrates how we sample the important light and Figure 3(c) shows an example of environment map sampling.

6 GPU Implementation of Caustics Rendering

We present the GPU implementation of our caustics rendering under D directional lights. Our implementation uses OpenGL and Cg as the rendering APIs. We use the image-based method proposed by Wyman et al. [9] to render the refraction effect of the caustic object.

Storing Caustic Spheres. We store the precomputed caustic spheres (in latitude-longitude format) of each light direction in a 3D texture format so that we can directly use the trilinear interpolation provided internally by the graphics API when we sample the caustic spheres. We can store the caustic spheres either in **multiple 3D textures** (each 3D texture stores the caustic spheres of one light direction) or a **single 3D texture** (which tiles the caustic spheres of all light directions into one 3D texture). By using multiple 3D textures, we need to do multiple rendering passes (D passes) in order to accumulate the caustic patterns from all light directions. On the other hand, by storing in a single 3D texture we are able to accumulate the caustic patterns from all D light directions either in single pass or multiple passes.

Computing Directional Light Radiance. The radiance for each light direction is computed by rendering the segmented regions of the environment map from the center of the caustic object. We render the segmented regions to a texture array whose resolution is $32 \times 32 \times D$, with each slice in the texture array corresponds to one segmented region. Afterward, we compute the total radiance by multiplying the number of pixels in that light region with the average radiance (computed by sampling the topmost level of the mipmap).

Caustics Sampling. There are two ways for computing the caustic intensity from all lights. First, for each light we sample the four nearest precomputed caustic spheres directly (**direct sampling**) as explained in Section 4. However, this becomes a bottleneck in volumetric caustics rendering since we need to do this process for every sampling point on the cast ray. Second, we compute and accumulate the caustic patterns of all lights beforehand (**compiled**) into a compiled 3D texture and then we just sample the compiled 3D texture.

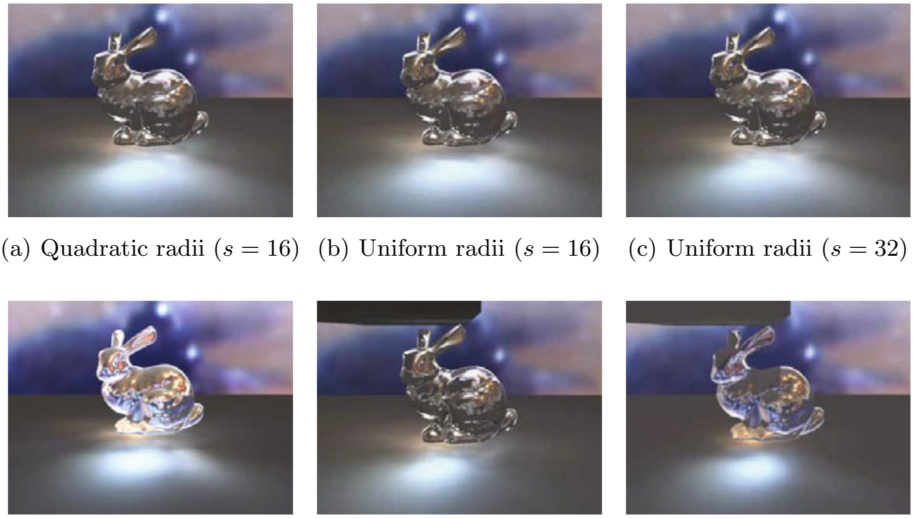
Rendering Passes. The accumulation of the caustics from all light directions can be computed either in **multiple passes** or **single passes**. In **multiple passes**, the iteration is in CPU with each iteration corresponds to one directional light and we may store the precomputed caustic patterns in the GPU either in multiple 3D textures or in a single 3D texture. For **single pass**, the iteration through all D directional lights is performed in GPU and the information of all directional lights are stored in Texture Buffer Object. For this technique, we can store the precomputed caustic spheres in GPU only in a single 3D texture.

7 Results

We performed the experiments on a PC with an Intel Core i7 2.67 GHz and an Nvidia GTX 285. The image size of our real-time rendering results are 1024×768 pixels.

Rendering Results. Figure 1 shows our rendering results under one directional light ($D = 1$ and $s = 16$, with D is the number of directional lights and s is the number of caustic spheres) and environment illumination ($D = 24$ and $s = 16$). From our experiments, the visual differences of the caustics between $D = \{32, 48\}$ and $D = 24$ were hardly noticeable. Moreover, the visual differences of $s = 16$ and $s = 32$ were also not very apparent. Therefore, we suggest the rendering with $D = 24$ and $s = 16$.

Sampling Comparison. Figure 4 shows the comparison of caustics rendering using our quadratic radii caustic spheres and uniform radii caustic spheres



(a) Quadratic radii ($s = 16$) (b) Uniform radii ($s = 16$) (c) Uniform radii ($s = 32$)

(d) mental ray (without occlusion) (e) Occlusion - Quadratic radii ($s = 16$) (f) mental ray (with occlusion)

Fig. 4. Comparisons of the rendering results. (a), (b), (c), and (e) were rendered using 24 directional lights.

proposed by Wyman et al. [8]. As seen in Figure 4, the quadratic radii using fewer caustic spheres achieves similar visual results as the uniform radii which uses more caustic spheres. With few caustic spheres ($s = 8, 16$) circular banding artifact is visible near the caustic object in the rendering results using uniform radii caustic spheres due to the insufficient number of caustic spheres near the caustic object.

Comparison with mental ray. We also compare our rendering results with the results rendered using mental ray (which took about four minutes to render each frame). Figures 4(a) and 4(d) show the comparison without light occlusion and Figures 4(e) and 4(f) show the comparison with light occlusion. Note the similarity between our results and mental ray results such as the caustics in front and behind the bunny (for the example without occlusion) and the vanished parts of the caustic patterns (for the example with occlusion).

Performance Comparison. We performed experiments using combinations of all possible options of the rendering techniques described in Section 6 and combinations of various numbers of directional lights $D = \{1, 16, 24, 32, 48\}$ and caustic spheres $s = \{8, 16, 32\}$ to determine the best rendering performance (in average frames per second, fps). In general, we achieved the best rendering performance under environment illumination for $D = 24$ and $s = 16$ (29.95 fps for cast caustics rendering and 14.57 fps for cast caustics and volumetric caustics rendering) by using the combination of multiple passes rendering, multiple 3D textures storage for the caustic spheres, and the compiled technique.

8 Conclusions and Future Work

We have presented a technique for real-time rendering of caustics and volumetric caustics in dynamic scenes under multiple directional lights. Our technique can be applied to render approximate caustics under environment illumination taking into account light occlusion from the surrounding objects. The limitations of the proposed technique are as follows. Our technique requires a large amount of memory to store the caustic spheres. Thus, we are interested in finding a method to compress the caustic spheres. Since we perform precomputation, the caustic objects are not deformable.

Acknowledgments. This work has been supported by the National Research Foundation grant, which is administered by the Media Development Authority Interactive Digital Media Programme Office, MDA (IDMPO). We would like to express our gratitude to Paul Debevec for the environment cube maps (Campus, Kitchen, and Grace Cathedral) and Stanford Computer Graphics Laboratory for the 3D models (bunny, dragon, armadillo).

References

1. Debevec, P.: A median cut algorithm for light probe sampling. In: SIGGRAPH 2005: ACM SIGGRAPH 2005 Posters, p. 66. ACM, New York (2005)
2. Günther, J., Wald, I., Slusallek, P.: Realtime caustics using distributed photon mapping. In: Keller, A., Jensen, H.W. (eds.) Rendering Techniques 2004: Eurographics Symposium on Rendering, pp. 111–121. Eurographics Association, Eurographics (2004)
3. Hadwiger, M., Ljung, P., Salama, C.R., Ropinski, T.: Advanced illumination techniques for GPU volume raycasting. In: SIGGRAPH Asia 2008: ACM SIGGRAPH Asia 2008 Courses, pp. 1–166. ACM, New York (2008)
4. Jensen, H.W.: Global illumination using photon maps. In: Proceedings of the Eurographics Workshop on Rendering Techniques 1996, pp. 21–30. Springer, London (1996)
5. Purcell, T.J., Donner, C., Cammarano, M., Jensen, H.W., Hanrahan, P.: Photon mapping on programmable graphics hardware. In: SIGGRAPH 2005: ACM SIGGRAPH 2005 Courses, p. 258. ACM, New York (2005)
6. Shah, M.A., Konttinen, J.: Caustics mapping: An imagespace technique for realtime caustics. IEEE Transactions on Visualization and Computer Graphics 13(2), 272–280 (2007)
7. Sun, X., Zhou, K., Stollnitz, E., Shi, J., Guo, B.: Interactive relighting of dynamic refractive objects. In: SIGGRAPH 2008: ACM SIGGRAPH 2008 Papers, pp. 1–9. ACM, New York (2008)
8. Wyman, C., Hansen, C., Shirley, P.: Interactive caustics using local precomputed irradiance. In: PG 2004: Proceedings of the Computer Graphics and Applications, 12th Pacific Conference, pp. 143–151. IEEE Computer Society, Washington (2004)

9. Wyman, C.: Interactive image-space refraction of nearby geometry. In: GRAPHITE 2005: Proceedings of the 3rd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia, pp. 205–211. ACM, New York (2005)
10. Wyman, C., Davis, S.: Interactive image-space techniques for approximating caustics. In: I3D 2006: Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games, pp. 153–160. ACM, New York (2006)
11. Zhou, K., Hou, Q., Wang, R., Guo, B.: Real-time KD-tree construction on graphics hardware. In: ACM SIGGRAPH Asia 2008 papers, pp. 1–11. ACM, Singapore (2008)

An Extraction Method of Lip Movement Images from Successive Image Frames in the Speech Activity Extraction Process

Eung-Kyeu Kim[1], Soo-Jong Lee[2], and Nohpill Park[3]

[1] Dept. of Infor. & Commu. Engineering, Hanbat Nat'l Univ., Korea
kimeung@gmail.com
[2] Automatic Speech Translation Research Team, ETRI, Korea
sjleetri@etri.re.kr
[3] Dept. of Computer Science, Oklahoma State University, USA
npark@cs.okstate.edu

Abstract. In this paper, we propose an extraction method of lip movement images from successive image frames and present the possibility to utilize lip movement images in the speech activity extraction process of speech recognition phase. The image frames are acquired from the PC image camera with the assumption that facial movement is limited during talking. First of all, one new lip movement image frame is generated with comparing two successive image frames each other. Second, the fine image noises are removed. Each fitness rate is calculated by comparing the lip feature data as objectly separated images. It is analyzed whether or not there is the lip movement image through verification to the objects and three images which have higher rates in their fitnesses. As a result of linking the speech & image processing system, the interworking rate shows 99.3% even in the various illumination environments. It was visually confirmed that lip movement images are tracked and can be utilized in speech activity extraction process.

Keywords: Lip Movement image, Image Frames, Acoustic Noises, Speech & Image processing system.

1 Introduction

In recent years, information technologies have spread rapidly due to the miniaturization and increased mobility of these information appliances. These changes have made it necessary to develop an speech interface technology that can effectively control these information appliances. The most difficult obstacle in the speech recognition phase, through technology which converts speech into text, is acoustic noises. The service environment of speech recognition is full of acoustic noises. It is no exaggeration to say so. The process of speech recognition is plagued by a large variety of acoustic noises including noises from its own drives of the appliances, network noises, and other environmental noises. Channel noises or stationary noises are almost completely eliminated, because their sizes and frequencies are easily identified by

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 317–325, 2010.

their consistency. The real challenge lies in the elimination of more dynamic noises, which are more difficult to be identified due to their irregular sizes and frequencies. Once the acoustic noises entered into the speech recognition phase, the noises are not removed and are the main cause for the low speech recognition rate. By the way, regardless of the ambient acoustic noises, images will be acquired and processed continually.

This paper is a part of the scheme to utilize lip movement image in order to prevent acoustic noises from being recognized as speech[1]-[2]. There is certainly lip movement whenever someone speaks, and any sound energy is extracted. Also, the sound energy which is not correspondingly associated with lip movement would be identified as noises. If lip movement is confirmed[3]-[4] in the speech activity extraction process of speech recognition phase, lip movement can efficiently prevent the acoustic noises from being classified as speech.

In this paper, we propose an extraction method of lip movement images from successive image frames in the speech activity extraction process[5] which is preprocessing phase of speech recognition. The image frames are acquired from the PC image camera. We also present the possibility to utilize lip movement images in the speech activity extraction process of speech recognition phase. Ultimately, we try to apply this extraction method to speech recognition of Robert surrounded with any outside environment including the dynamic acoustic noises.

This paper is organized as follows. Section 2 describes image frame acquisition environments. In section 3, the procedure of lip movement image extraction is described including the noise image deletion and the extraction of lip movement image features. In section 4, we show the verification of lip movement images and it's experimental environment, where the template matching method is showed to verify lip movement images accurately. In section 5, we bring to conclusions.

2 Image Frame Acquisition Environments

As the multimedia environment has progressed rapidly, cameras are equipped with all kinds of information appliances, which can easily, acquire movement images, and process them. Especially image cameras for PCs have come into use widely in the recent years and those allowed the general public to generate images directly, and to edit them. The PC image camera is mainly used for image communication with a remote site or image chatting, but it can also be utilized continuously in acquiring, comparing and analyzing the image frames. Nowadays, the PC image camera generates 320*240 sized color images at the minimum speed of more than 15 frames per second. Thus it is sufficient to study the process of acquiring and processing image frames.

Fig. 1 shows speech and image processing system in the speech recognition phase. The processing system is composed of three parts, that is, image processing part, and speech processing recognition part, and shared memory part. Lip movement images can be acquired under the lighting conditions of the general home and office environment. It is also assumed that there was no excessive movement of the head or face during the conversation. Therefore, excessive movements of a face were excluded from this study.

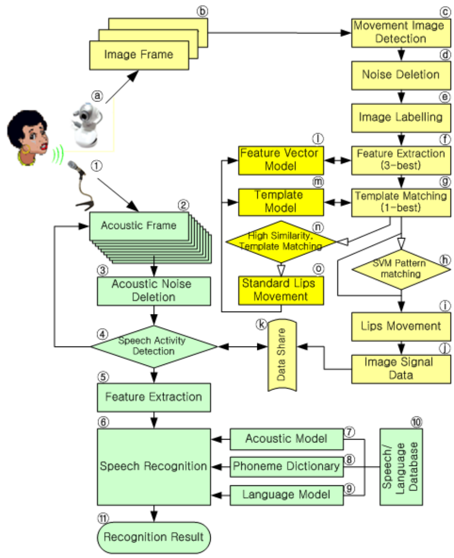


Fig. 1. Speech and image processing system in the speech recognition phase

3 Procedure of Lip Movement Image Extraction

The extraction of lip movement images uses a five phase processes, which is as follows: acquiring movement images, removing fine image noises, separating the partial images as different areas, comparing the area images to lip features, and verifying whether it is the lip image or not, etc. This process is repeated for each image frame unit.

Figure 2 shows five steps of the process for extracting lip movement image. Image frames "a" and "b" are the shape of the facial components when they are changing in the speaking situation.

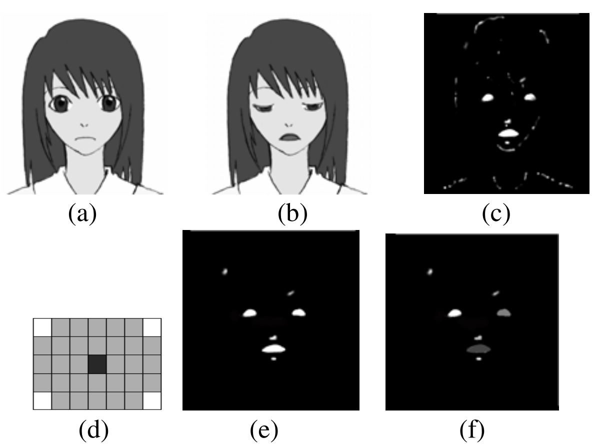


Fig. 2. Process for extracting lip movement image

"c" is composed of the results which are changed between "a" and "b". As a structuring element "d" is used to remove minute image noises which appear in "c". "e" is the result of the application of the structuring element. "f" is the result of the different image areas, which have different pixel value. The following sections address the specific processes involved in the completion of component "f".

3.1 Movement Distinction and It's Image Acquisition

In order to distinguish whether or not there is a lip movement, a comparison between one image frame and the other image frame must be done in a pixel unit. A fine difference of pixel value in each frame is constantly occurring phenomenon because of the image processes and illumination. Therefore, after setting up a threshold for a pixel value difference, it is desirable to distinguish those values which are in excess of the threshold. The setting of a threshold value is necessary to consider environmental illumination and PC computing performance. Here the threshold was set to 30. Therefore, the decrease of the computational scale was processed after the color images were converted to monochrome.

In order to acquire and compare the successive image frames, a minimum of two image buffers are necessary. While continuously updating the new images, they are simultaneously compared to the buffer images by pixel unit each other. Whether or not there are image frames or their components change is interpreted according to the differences in the related pixel value. The results which are compared with a pixel unit are acquired, and those are composed of new movement image frames.

3.2 Image Noise Removal and Movement Image Separation

The fine image noises are removed by the "opening"[6] technique as one of the morphological image processing method. The "opening" technique is composed of a two step technique, "erosion" and "dilation". By applying the structuring element of Figure 2(d), to images smaller than the structuring element, the smaller images are removed. The structuring element used in "opening" can be set up with several directions, that is, up and down, right and left, diagonal, and so on. This process is expressed by their symbols.

$$e = c \circ d = (c \,\Theta\, d) \oplus d$$

The structuring element ("d") is set up as 7*5 pixel size with a length that is longer than the width that takes into consideration the lip movement image shape. The lip movement image width is shorter than the length.

Nevertheless after removing the detailed image noises, many other large sized movement image parts will remain. These image parts include eye blink unrelated to eye movements, jaw movements, face shake, and many extra movements in the background and the light. In addition, there are lots of extra images according to the background and the light.

In order to isolate the lip movement image, it is necessary to separate those residual image parts into different areas each other. The separation of the image part is performed by the "grassfire transform"[7] technique. Through all the image pixels, it makes groupings of the neighborhood pixels with the same value by means of giving a distinct value to each.

3.3 Application of Lip Movement Image Features

After the objects are separated into different areas, a variety of data can be extracted, such as width, lengthwise, pixel numbers, position, etc. If a comparative analysis of the data is done, the characteristics of lip image can be determined.

Several features of lip movement images can be determined mainly through it's relation to the eyes. First, it is the extent of the rectangular form. Second, it is done by finding the center of the image in relation to the eyes and the lip. Third, it is the number of pixels. Although the pixel numbers vary largely according to the lip movement scales, the maximum and/or minimum counts are important to distinguish each image area.

Table 1 below is the results of the data for the feature elements of lip movement images which are collected through a PC image camera at the interval of 50cm.

The dimension rate is calculated as the actual pixel numbers divided by the rectangular dimension's maximum size for width and length.

The fitness rate of the lip movement feature is calculated by combing all the feature data above. The more the fitness rate of the image is high, the more the possibility of the lip movement image part is great. The next section will show the method of verification by identifying the three images which have the highest fitness rates.

Table 1. Feature elements of lip movement images

	width	legnth	width/leng.	dimension rate	pixels	length location	width location
collected data (distance:50cm)	6	24	3.630	0.436	330	0.733	0.812
	5	28	4.140	0.426	329	0.873	1.869
	3	23	4.800	0.322	214	0.707	0.932
	4	23	5.400	0.335	275	0.727	0.935
	6	28	4.667	0.438	384	0.669	0.947
	5	24	4.800	0.360	333	0.736	0.994
	8	31	3.875	0.492	504	0.707	1.424
	4	18	4.500	0.346	208	0.747	1.194
	7	19	2.714	0.652	204	0.480	1.011
	3	4	3.667	0.327	101	0.600	2.404
	4	17	4.500	0.393	183	1.013	0.980
	2	20	3.000	0.324	37	0.667	1.014
	6	18	4.667	0.403	417	1.136	0.938
	3	19	3.333	0.337	89	0.693	1.392
	……	……	……	……	……	……	……
average	4.6	18.1	4.284	0.395	251	0.781	1.282
standard dev.	1.453	7.053	0.776	0.070	121	0.180	0.399
maximum	8	31	6.250	0.652	504	1.175	2.404
minimum	2	4	2.714	0.286	37	0.480	0.812

4 Verification of Lip Movement Images and Experimental Environments

The preliminary step in the verification process should be the establishment of threshold values for the fitness rates of a lip movement image. However, the fitness levels for eyes and lip movement images are changeable on occasion.

The fitness rates are made as the result of the correlation rate from the template matching, which is added to the fitness rate by feature elements. The fitness rates are used as a final score to decide the lip movement image. The following Figure 3 shows the partial results that were collected from the largest verification rate per every frame.

The horizontal axis represents the fitness rates and vertical axis represents the frequencies. The highest convex curve is related to the lip image areas, and the second convex curve is a related to the eyes. The valley point between the first and the second convex curve is a threshold classifying the lip image area and the others. The threshold value is formed and extracted automatically. The threshold value is situated at the point between two convex curve changes. where, valley point is converged into a unified value, namely 0.57.

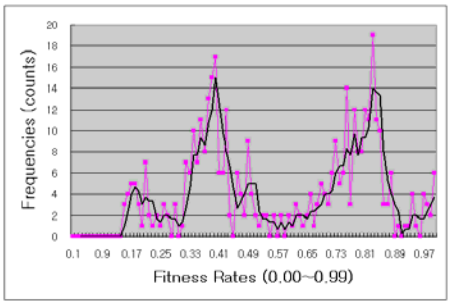
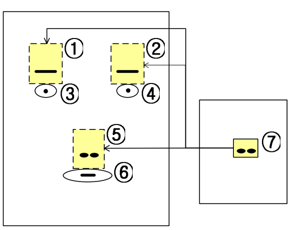


Fig. 3. Fitness rates and frequencies

The next step is considering the template matching technique which can be used for accurate verification. Template matching is a suitable method for measuring the correlation between two images in a pixel unit. The correlation can compute the differences between the two images per related pixels. Fig. 4 shows the model (a) and the calculating formula (b) for measuring the template matching rate.



(a)

$$r = \frac{a.b}{|a|.|b|} = \frac{\sum_{x=0}^{M}\sum_{y=0}^{N} a(x,y)b(x,y)}{\sqrt{\sum_{x=0}^{M}\sum_{y=0}^{N} a(x,y)^2 \sum_{x=0}^{M}\sum_{y=0}^{N} b(x,y)^2}}$$

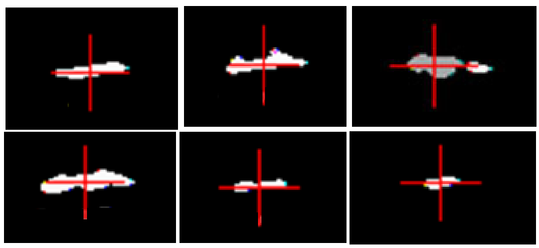
(b)

Fig. 4. Template matching rate measuring model (a), calculating formula(b)

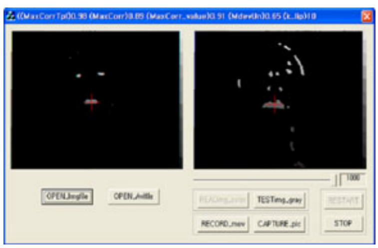
In Fig. 4(b), a(x,y) shows the brightness value considering the average value of E(g) obtained from the input image g(x,y). Additionally, b(x,y) shows the brightness value considering the average value of E(t) obtained from the template image t(x,y). During the implementation of the template matching technique, only three movement images which are selected in order of their fitness levels are actually verified.

The lip movement itself should not be included as an element in the template during talking. Unlike the lips, the nose is unchangeable even during talking, so the difference between the light and dark shade is distinct.

Figure 5(a) shows tracking results of three lip movement image candidates and 5(b) does tracking of a final lip movement image which is selected from three lip movement image candidates.



(a)



(b)

Fig. 5. (a) Trackings of three lip movement image candidates; (b) Tracking of a final lip movement image

Prior to the combined speech and image processing system, it was tested whether the speech recognition phase is influenced by external acoustic noise which is not a part of the speech recognition object. The confirmation of illumination environment was investigated while the brightness of the lighting application was changed. Additionally, it was determined whether any acoustic noise was blocked during the speech recognition phase through the combination of the speech and image processing system. These experimental results are summarized in table 2, and table 3. This system was implemented by linking the existing speech recognition engine with the image processing test bed for lip movement image tracking. The computer utilized for the experiment was a PC Pentium IV with a 3.6GHz processor under the general office environment.

Table 2. The image processing and illumination adaptation

Tracking object	Illumination level (lx)	Success rate (%)
Lip movement	Office (500~300) Home (300~100) Laboratory (100~)	About 95

Table 3. Combined speech and image system

Input object	Content	Success rate (%)
Acoustic noise	Blocking of the speech recognition progress	100
Speech utterance	Speech recognition execution	99.3
	Speech recognition non-execution	0.7

5 Conclusion

In this paper, we proposed a method to extract lip movement images from successive image frames and presented the possibility to utilize lip movement images in the speech activity extraction process of speech recognition phase. The image frames are acquired from the PC image camera with the assumption that facial movement is limited during talking. First of all, one new lip movement image frame is generated with comparing two successive image frames each other. Second, the fine image noises are removed. Each fitness rate is calculated by comparing the lip feature data as objectly separated images. It is analyzed whether or not there is the lip movement image through the verification to the objects and three image candidates which have higher rates in their fitnesses. As a result of linking the speech and image processing system, the interworking rate showed 99.3% even in the various illumination environments. It was visually confirmed that lip movement images are tracked and can be utilized in speech activity extraction process. Ultimately, we try to apply this extraction method to speech recognition of Robert surrounded with any outside environments including the dynamic acoustic noises. In the future study, the confirmation of the real-time processing for extracting further robust lip movement images remains to be solved.

References

1. Chan, M.T., Zhang, Y., Huang, T.S.: Real-Time Lip Tracking and Bimodal Continuous Speech Recognition. In: IEEE Second Workshop on Multimedia Signal Proceeding, December 7-9, pp. 65–70 (1998)

2. Potamianos, G., Neti, C., Luettin, J., Matthews, I.: Audio-visual automatic speech recognition: An overview. In: Issues in Visual Speech Processing. MIT Press, Cambridge (2004)
3. Liew, A.W., Leung, S.H., Lau, W.H.: Lip contour extraction from color images using a deformable model. Pattern Recognition 35(12), 2949–2962 (2002)
4. Liew, A., Wang, S. (eds.): Visual Speech Recognition: Lip Segmentation and Mapping. IGI Global (2009)
5. Libal, V., Connell, J., Potamianos, G., Marcheret, E.: An embedded system for in-vehicle visual speech activity extraction. In: Proceedings of the International Workshop on Multimedia Signal Processing (MMSP 2007), Chania, Greece, pp. 255–258 (2007)
6. Wu, Z.Q., Ware, J.A., Stewart, W.R., Jiang, J.: The Removal of Activitying Effects Caused by Partially Overlapped Sub-activity Contrast Enhancement. Journal of Electronic Imaging 14(3), 033006(8 pages) (2005)
7. Leymarie, F., Levine, M.D.: Simulating the Grassfire Transform Using and Active Contour Model. Trans. IEEE Pattern Analysis and Machine Intelligence 14(1), 56–75 (1992)

Rule-Based Camerawork Controller for Automatic Comic Generation from Game Log

Ruck Thawonmas, Ko Oda, and Tomonori Shuda

Intelligent Computer Entertainment Laboratory
Graduate School of Science and Engineering, Ritsumeikan University
Kusatsu, Shiga, 525-8577, Japan
ruck@ci.ritsumei.ac.jp

Abstract. We propose a rule-based camerawork controller for a recently proposed a comic generation system. Five camerawork rules are derived through an analysis of online-game webcomics about Lineage 2, one rule for each of the five event types: chatting, fighting, moving, approaching, and special. Each rule consists of three parts relating to the three camera parameters: camera angle, camera position, and zoom position. Each camera-parameter part contains multiples shot types whose value indicates the frequency of their usages in the analyzed webcomics. In this paper, comic frames generated with the proposed camerawork controller are shown and compared with those generated with our previous controller based on heuristic rules, confirming the effectiveness of the proposed camerawork controller.

1 Introduction

Comic is a promising media for summarizing experiences in an entertaining fashion. Recent comic applications include summarization of activities in a conference [1], daily activities [2], video sequences [3], and game-play activities [4–6]. Comic-based experience summaries facilitate augmenting personal memories as well as promoting communication among user communities.

Since 2008, we have developed a number of techniques [7,8,9] for our comic generation system [5] that aims at automatically generating comic from game log. In our system, we adopted the same approach as in [4] where the game engine of a game of interest is used for rendering comic based on information in the game log. Another approach adopted in [6] is that of composing comic from selected screen shots. Although the game engine must be accessible, the former approach provides more room to play with camerawork and thus more varieties in comic.

Recently, in order to utilize the advantage of the game-engine approach, we proposed in [9] a module for manually editing the camerawork of generated comic. However, its initial camerawork, heuristically decided by the authors, lacks varieties and requires a large amount of editing work and thus causes a high burden to the user. To ease this burden, we address here the issue on automatic camerawork control and propose a camerawork controller that automatically

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 326–333, 2010.

Fig. 1. Screenshot of The ICE

decides the camerawork of each frame based on rules derived from an analysis of online-game webcomics.

The contributions of this paper are as follows:

1. Universal and reliable camerawork rules derived from webcomics about Lineage 2,
2. The proposed camerawork controller utilizing the derived rules and providing rich varieties of shots,
3. Outline of our automatic comic generation system applicable to any game, provided that its game engine is accessible.

As with our recent papers, we use, as the research platform, an online game called The ICE (Fig. 1), under development at the authors' laboratory. In The ICE, typical online-game activities, such as monster fighting, chatting with other characters, and item trading, are available.

2 Comic Analysis

We selected the web comics [10] in Japanese whose stories are about Lineage 2 for analyzing camerawork rules. Our reasons for this selection, besides the third author himself being a resident of Lineage 2 and thus familiar with the contents, are as follows:

Table 1. Resulting camerawork rule for each event type

Event Type	LA	EL	HL	BE	F	TF	S	TB	B	SC	C	M	K	L	SL
Chatting	266	5807	344	74	2847	2489	512	242	255	432	1265	2939	866	770	136
Fighting	65	631	49	0	276	265	66	48	65	40	74	281	162	147	33
Moving	4	97	13	0	35	31	26	5	15	5	13	30	18	33	13
Approaching	2	21	1	0	19	1	2	0	1	0	5	9	4	4	2
Special	122	1733	199	74	962	779	140	58	110	188	357	897	341	243	44

– Their stories are based on Lineage 2, fitting to our comic genre.
– Each story has a sufficient length, ensuring the reliability of analysis results.
– Multiple comic writers contribute to the above site, preventing the analysis results from being biased by a single comic writer.

In order to increase the quality and reliability in analysis, we asked two college students to independently read and analyze, based on the same analysis manual, all comics in this site. Their task was to find the frequency of each shot type relating to the camera angle, camera position, and zoom position in the comic frames for five event types defined as follows:

Chatting event that includes at least one dialogue or chat balloon
Fighting event that includes a content relating to fighting activities
Moving event that shows movement of characters
Approaching event where at least two characters or objects are approaching each other
Special event to which the above four events do not apply.

Note that multiple event types might be applicable to an analyzed comic frame. For each of such frames, the shot types for the camera angle, camera position, and zoom position were shared by and counted for all applicable event types.

Table 1 shows the analysis results integrating the individual results from the two analyzers. An element in this table indicates the frequency of a shot type of interest in the corresponding event. The number of shot types relating to the camera angle, camera position, and zoom position are four, five, and six, respectively; i.e.,

Camera angle: Low Angle, Eye Level, High Angle, and Bird Eye shots;
Camera position: Front, Tilted Front, Side, Tilted Back, and Back shots;
Zoom position: Super Close-up, Close-up, Medium, Knee, Long, and Super Long shots;

where the shot target is the last character who chatted and who performed a fighting action for the chatting event and the fighting event, respectively; for the other event types, the shot target is the player character.

For each shot type of a camera parameter of interest, we use the frequency in Table 1 as its weight in the corresponding event type. The camerawork controller, described in the next section, uses these weights in a roulette-wheel fashion for

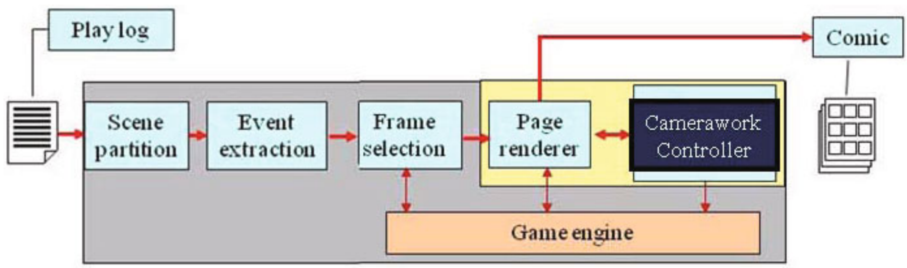


Fig. 2. Architecture of the comic generation system

deciding a shot-type of each camera parameter for a given event type. For example, considering selection of a shot type of the camera angle for the fighting event, the roulette wheel has four areas with the sizes in the proportion of 65:631:49:0.

3 Camerawork Controller and Comic Generation System

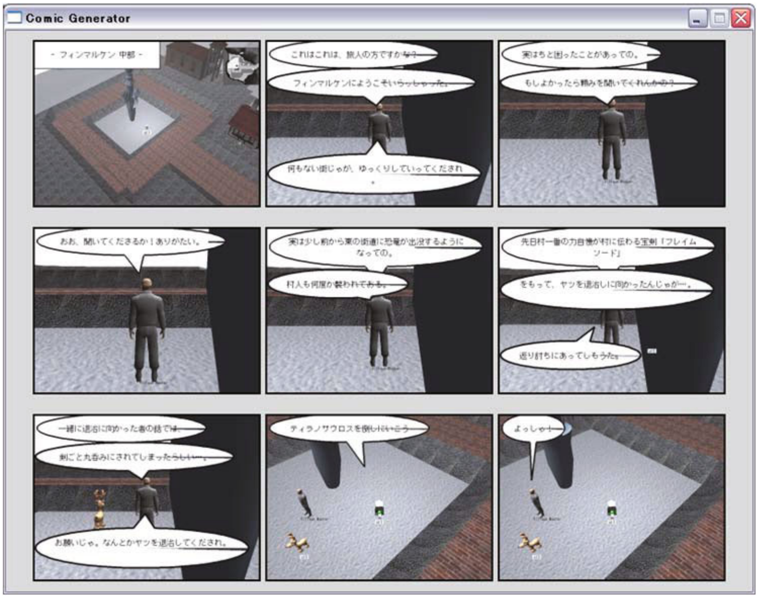
Figure 2 depicts an overview of our comic generation system [5] that includes the proposed camerawork controller as the most right module. In this system, first, the play log is chronologically divided into multiple partitions, each called a scene, connecting a story (if any) in the play. At the event extraction module, a scene is further divided into multiple chunks, each called an event composed of multiple actions such as fighting, chatting, and moving. After frames have been selected from those events at the frame selection module [8], the parameters for deciding frame size and shape are decided. These parameters are used together with the camera angle, camera position, and zoom position for rendering comic frames.

The proposed camerawork controller decides the camerawork parameters for a given frame through the following three steps:

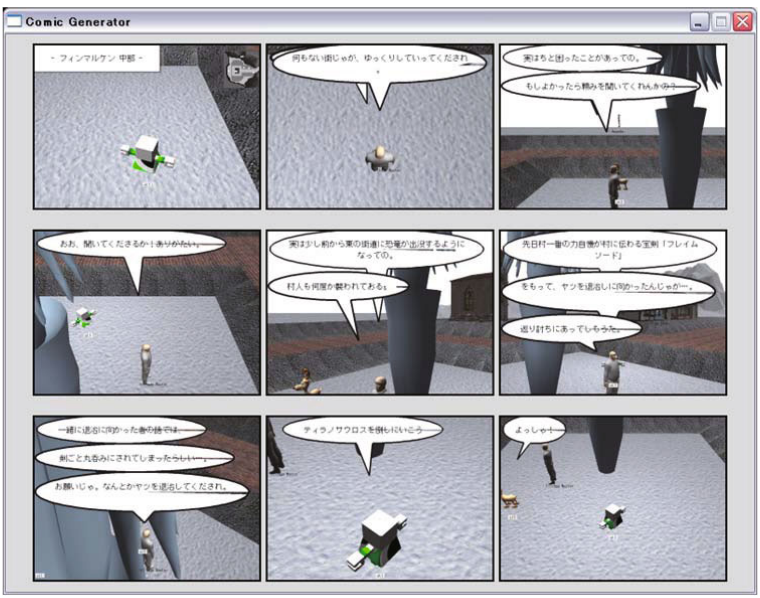
1. Determine the event type by using the majority voting of all action types occurring in the frame
2. Perform roulette-wheel selection of a shot type of each camera parameter for the determined event type
3. Set the value of each camera parameter to the predefined value of the selected shot type.

4 Results and Discussions

Figures 3.a and 3.b show a comic page generated with the previous camerawork controller, used in 5,7–9, and that with the proposed camerawork controller, respectively. This is from a scene where the player, the robot-like character, and his friend, the dog-like character, are talking to a mission master, the human-like character, in order to receive a game mission. Because the former controller uses

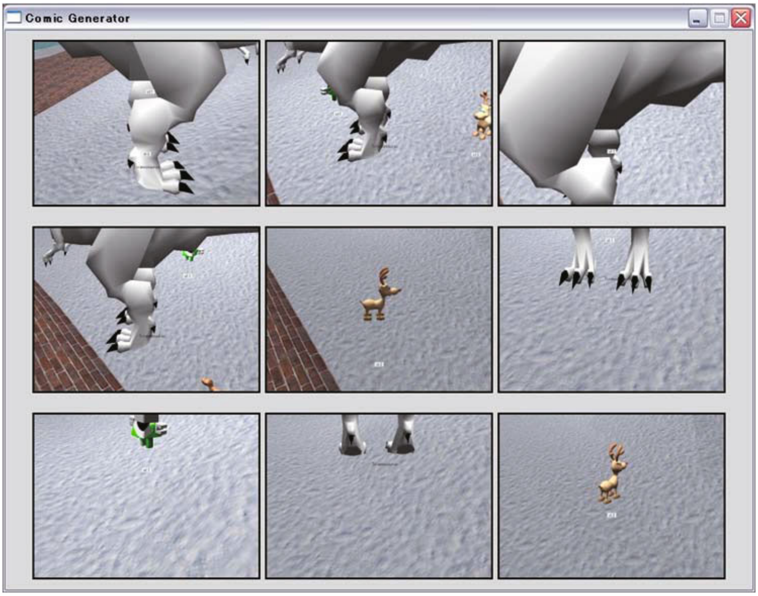


(a) Previous

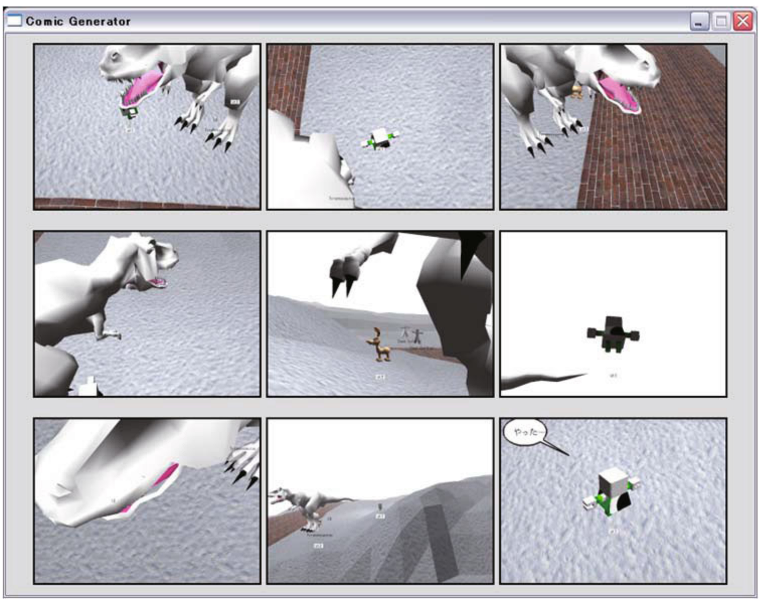


(b) Proposed

Fig. 3. Comic page for a chatting scene with the previous camerawork controller and the proposed camerawork controller



(a) Previous



(b) Proposed

Fig. 4. Comic page for a fighting scene with the previous camerawork controller and the proposed camerawork controller

only one combination of the camera angle, camera position, and zoom position for each event type, there exist two frame subsequences, each having a same shot type, in the former figure. On the contrary, the latter figure has a richer variety of shot types. Visual comparison between these figures and that between Figs. 4.a and 4.b for a fighting scene also confirm the superiority of the proposed camerawork controller.

5 Conclusions and Future Work

This paper described our camerawork controller that decides a combination of the camera angle, camera position, and zoom position for a frame according to the frame's event type. Such a decision is done for a given event type using roulette-wheel selection where each shot-type's weight of a camera angle of interest represents its portion in the wheel. An analysis of Lineage 2 webcomics was conducted to derive those weights. Visual comparison between our previous camerawork controller and the proposed camerawork controller confirms the superiority of the latter.

Our future work includes fine tuning of the camera angle, camera position, and zoom position based on the entropy of each frame segment formed with a technique such as the Berkeley Segmentation Engine [11].

Acknowledgements

This work was supported in part by Grant-in-Aid for Scientific Research (C), No. 20500146, the Japan Society for Promotion of Science, and by Global COE (Center of Exellence) Progam "Digital Humanities Center for Japanese Arts and Culture".

References

1. Sumi, Y., Sakamoto, R., Nakao, K., Mase, K.: ComicDiary: Representing Individual Experiences in a Comic Style. In: Borriello, G., Holmquist, L.E. (eds.) UbiComp 2002. LNCS, vol. 2498, pp. 16–32. Springer, Heidelberg (2002)
2. Cho, S.B., Kim, K.J., Hwang, K.S.: Generating Cartoon-Style Summary of Daily Life with Multimedia Mobile Devices. In: Okuno, H.G., Ali, M. (eds.) IEA/AIE 2007. LNCS (LNAI), vol. 4570, pp. 135–144. Springer, Heidelberg (2007)
3. Calic, J., Gibson, D.P., Campbell, N.W.: Efficient Layout of Comic-like Video Summaries. IEEE Transactions on Circuits and Systems for Video Technology 17(7), 931–936 (2007)
4. Shamir, A., Rubinstein, M., Levinboim, T.: Generating Comics from 3D Interactive Computer Graphics. IEEE Computer Graphics and Applications 26(3), 53–61 (2006)
5. Shuda, T., Thawonmas, R.: Automatic Comic Generation Using Online-Game Play Log. Journal of Game Amusement Society 3(1), 40–45 (2009) (in Japanese)

6. Chan, C.-J., Thawonmas, R., Chen, K.-T.: Automatic Storytelling in Comics: A Case Study on World of Warcraft. In: CHI Extended Abstracts 2009, pp. 3589–3594 (2009)
7. Thawonmas, R., Shuda, T.: Comic Layout for Automatic Comic Generation from Game Log. In: Proc. IFIP International Federation for Information Processing (ECS-2008), vol. 279, pp. 105–115 (2008)
8. Shuda, T., Thawonmas, R.: Frame Selection for Automatic Comic Generation from Game Log. In: Stevens, S.M., Saldamarco, S.J. (eds.) ICEC 2008. LNCS, vol. 5309, pp. 179–184. Springer, Heidelberg (2008)
9. Thawonmas, R., Oda, K., Shuda, T.: Camerawork Editor for Automatic Comic Generation from Game Log. In: CD-ROM Proc. of Nicograph International 2009, Kanazawa, Japan, June 19-20 (2009)
10. Lineage 2 webcomics: http://lineage2.plaync.jp/l2fun/comic/menu.aspx (last accessed on June 22, 2010)
11. Martin, D., Fowlkes, C., Malik, J.: Learning to Detect Natural Image Boundaries Using Local Brightness, Color and Texture Cues. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(5), 530–549 (2004)

A Framework for Constructing Entertainment Contents Using Flash and Wearable Sensors

Tsutomu Terada[1] and Kohei Tanaka[2]

[1] Kobe University, 1-1 Rokkodai, Nada Ward,
Kobe, Hyogo 657-8501, Japan
tsutomu@eedept.kobe-u.ac.jp
[2] Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan
tanaka.kohei@ist.osaka-u.ac.jp

Abstract. Multimedia interactive contents that can be controlled by user's motion attract a great deal of attention especially in entertainment such as gesture-based games. A system that provides such interactive contents detects the human motions using several body-worn sensors. To develop such a system, the contents creator must have enough knowledge about various sensors. In addition, since sensors and contents are deeply associated in contents, it is difficult to change/add sensors for such contents. In this paper, we propose a framework that helps contents creators who do not have enough knowledge on sensors. In our framework, an interactive content is divided into two layers; sensor management layer and content layer. We confirmed that creators can create interactive contents easier with our framework.

Keywords: Wearable Computing, Flash, Development Environments.

1 Introduction

In recent years, the importance of intuitive human-computer interfaces has increased and increased. Especially, *Nintendo Wii* that is a video game console released in 2006 infuses new breath into the game market since we can play video games intuitively by using sensor-enabled game controller as a sword and a tennis racket. On the other hand, it is still difficult for contents creators who create contents using *Flash* or other animation tools to develop such contents since they usually do not have the knowledge to manage sensors and to recognize user's motion from raw sensor data. Moreover, once the contents creator made a content using sensors, it is difficult to change the sensors used in contents because the sensor management is deeply integrated to the contents. Therefore, there is a requirement on the support for contents creation by easier sensors managements.

Here, Flash[6] is so popular for providing interactive contents on web browsers. Flash contents have various advantages such as small file size with respect to rich contents, and support for various platforms (more than 99% people in mature market can play flash contents). Contents creators can implements

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 334–341, 2010.

```
WHEN GPS_MOVE
IF CURRENT.Position == 'station'
THEN DO BROWSER_OPEN('URL')
```

Fig. 1. An example of ECA rule

interactivity on the contents using easy programming scripts called *ActionScript*. Thus, we propose a framework for flash creators to create interactive contents with sensors and actuators. Our framework provides several tools to hide the difficulty in sensor management and creators can create interactive contents by ActionsScript easily.

The remainder of this paper is organized as the follows. Section 2 describes our environmental assumptions and Section 3 presents related works. Section 4 describes the process of developping contents using our framework. We explain our framework in detail in Section 5 and discuss our framework in Section 6. Finally, we conclude this paper in Section 7.

2 Environmental Assumption

The target users of our framework are flash contents creators that do not have enough knowledge how to use the sensor data while they can create flash contents without sensors. Recent days, there are many people that can create flash contents since flash becomes popular.

We employ *Wearable toolkit* that we developed to define/recognize user contexts. It consists of an event-driven rule processing engine and several plug-ins to implement context-aware applications easily. The rule in the toolkit is called ECA rule, which consists of *Event*, *Condition*, and *Action*. For example, when we want to develop the system that *displays the train schedule when user arrives at the station*, we define the reception of GPS data as *Event*, the comparison the current position with the position of station as *Condition*, and the navigation to web site as *Action*, as shown in Figure 1. Wearable toolkit has *Context Definition Tool* that users define user contexts easily by actually demonstrating the motion to be registered. The rule engine raises an event when it detects the similar motions to registered contexts. The Context Definition Tool supports various sensors such as 3-axis acceleration sensors, temperature sensors, a GPS, and RFID tag readers.

Figure 2 shows an example of our assumed content and user. The user interacts with the content by his motions in front of the screen. He receives feedbacks by the screen, sound, and some actuators.

3 Related Works

There are many researches to develop services using various sensors and actuators. Phigets[3] and Teleo[2] are toolkits that consist of various sensors and

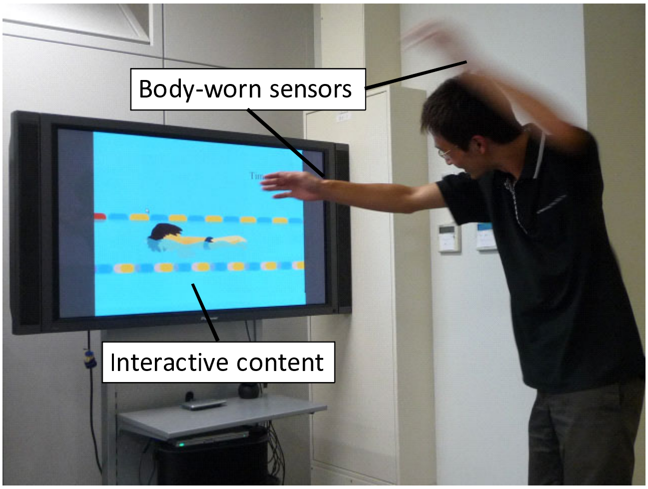


Fig. 2. An example of playing contents

actuators to develop applications using the libraries in various development environments. Gainer[8] is a hardware platform to connect various devices to computer. GlovePIE[1] is a tool for assigning device functions to application functions by a script language. Using these toolkits, we can develop applications that use various devices easily. However, we still need to know how to process the sensor data since they support only connection between computer and devices.

A CAPpella[7] enables us to define user actions easily, and GT2K[9], AR Toolkit[11], and DART[10] provide the recognition of gestures and markers by camera. However, these systems does not refer to programming model.

4 Contents Development with Our Framework

We show an actual development of interactive game, a 110m hurdles game, as the following steps. Our framework especially supports Step 2, 3, and 4 since these steps are difficult for contents creators.

Step 1. Contents design
Step 2. Motion registratoin
Step 3. Rules description
Step 4. Flash contents creation
Step 5. Debug

Step 1 Contents design: We established the overall picture of the contents in this step. Then, We discussed the purpose of the contents, motions to be used, and required sensors and actuators. In this case, the purpose of the game is to compete the time from start to goal. The character on the flash contents runs while the player runs, it stops when the player stops, and it jumps when the player jumps. Moreover, a vibrator works when the character fails to jump a hurdle.

Step 2 Motion registratoin: Next, we registered the player's motions that are designed in previous step using Context Definition Tool. To register them, we only performed the actual motions. We also confirmed that the actions can be recognized accurately. For this game, we registered the 3 motions; *run*, *stop*, and *jump*.

Step 3 Rules description: In this step, we described the rules for Wearable Toolkit to communicate with flash contents. Our *Code Generation Tool*, which is explained in the following section, produces the rules by inputting the relationship between user actions and flash scripts that are activated when the motions are recognized. Concretely, to create the rules about *run* action, we input the name of running action and a script that is executed at running. The Code Generation Tool generates the scripts and the rules to execute the input scripts when the system recognizes *run*. We also input descriptions about *stop* and *jump*. In the same way, we also easily get the scripts and the rules to control actuators by Code Generation Tool. In concrete, we selected the vibrating function from the list of Wearable Toolkit functions, then the named of the function can be used in flash contents.

Step 4 Flash contents creation: In this step, we created flash animation contents. The Code Generation Tool generated flash project file in which the script for communication is already described. To control an actuator, we only describe RaiseEvent() function that has the event name selected in the previous step. We added RaiseEvent() function in the script code for collision detection.

Step 5 Debug: We tested the contents using the debugger of Wearable Toolkit. Using this debugger, we confirmed that flash received the player motion events and Wearable Toolkit received the device control events. Moreover, we checked the accuracy in motion recognition. When the accuracy is not enough, we change the threshold in recognition or redefine the motions. We can change the content easily since Code Generation Tool generates rules and scripts independently from actions and actuators.

5 Design of Framework

Our framework consists of 2 layers as shown in Figure 3. *Sensor management layer* manages sensors and actuators, and recognizes user motions. *Flash contents layer* is flash contents including several templates to communicate with the other layer. Since it is difficult for flash creator to use the sensor data directly, our framework hides the sensor management layer from contents creators. If they define the user motions that they want to use semi-automatically using *Context Definition Tool*, flash contents can receive the motion-recognition event from sensor management layer.

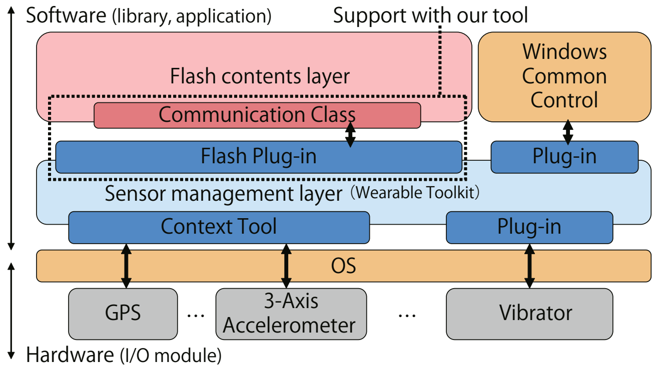


Fig. 3. Structure of proposed framework

In this research, we add a function for communicating with flash contents to Wearable toolkit and implement *Code Generation Tool* that generates the rules and the scripts for associating motions with actions.

To realize the communication between Wearable Toolkit and flash contents, we utilize the Flash Player's commands that are originally used for communication with Flash Player and the contents. Although this command cannot transfer without text message, it has enough capability to send context information to flash contents and we do not suppose the contents creators use the raw data from sensors. The communication between Wearable toolkit and flash occurs in the following three situations:

Communication on flash contents starting: First communication occurs in initializing the content. When a flash content starts, it sends messages to Wearable Toolkit to register user motions and actuators to be used via fscommand() function[4]. This process is based on the code shown in Figure 4, which is generated by Code Generation Tool. This process includes the association between motions and scripts. In the figure, `ACTION_NAME` action is associated with

```
RegisterAction(
    "ACTION_NAME", OnAction,
     "explanation", "type of argument");
RegisterEvent(
    "EVENT_NAME", "explanation",
     "type of argument");

function OnAction(Arg:type of argument){
 //OnAction implementation Here
}
```

Fig. 4. An example of script at initialization

```
WHEN CONTEXT_RECOGNIZED('CONTEXT_NAME')
IF CURRENT.Rate > 80
THEN DO FLASH_ACTION('ACTION_NAME')
```

Fig. 5. An example of ECA rule to send a message to the flash content

```
WHEN FLASH_EVENT('EVENT_NAME')
THEN DO CONTOL_ACTUATOR()
```

Fig. 6. An example of ECA rule to control actuators

OnAction() function. When the toolkit receives the messages, it registers the messages as events or actions.

Communication when user motion is recognized: After the flash content starts, Wearable Toolkit recognizes user motions from sensor data. When a motion is recognized, ECA rules that are generated by Code Generation Tool are executed to send the recognition results to the flash content. The flash content calls the associated scripts according to the received results.

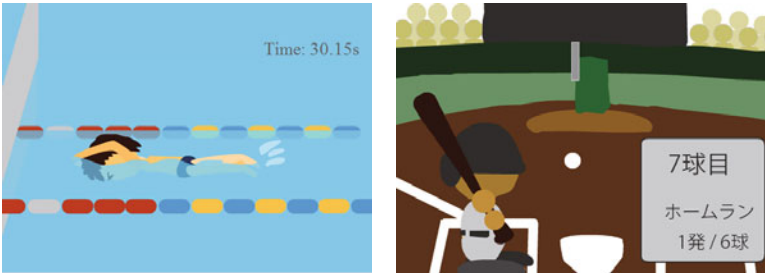
An example of such ECA rule is shown in Figure 5. This rule is activated when the context named `CONTEXT_NAME` is recognized, and sends the function name `ACTION_NAME` to exectute the script in the flash content. This rule is automatically generated by Code Generation Tool.

To realize this communication, `FLASH_ACTION` action calls SetVariable() function [5] on the SWFobject, which is a function to change the value of variables on the flash content. In the content, the target scripts are described as a callback function. In the example of Figure 4, `OnAction()` is executed when the value of the variable is changed to `ACTION_NAME`.

Communication when the content controls actuators: Our framework enables flash contents to control actuators by fscommand() function in the contents that sends text messages to Wearable toolkit. The ECA rules for actuators control in Wearable toolkit, which is created by our Code Generation Tool, receive commands and control the associated actuators. An example of the ECA rules is shown in Figure 6. The rule in the figure execute `CONTROL_ACTUATOR()` action to control an actuator when the command named `EVENT_NAME` is received via fscommand() from the flash content.

6 Consideration on Implementation Time

We discuss the easiness of contents development by using our framework. Firstly, we implemented an interactive game by modifying existing breakout game. We associated two motions on the game with two physical motions. We took less than one hour to modify this game.



(a) Swimming game (b) home run chase game

Fig. 7. Created games

Next, we implemented two games that are the homerun chase game and the swimming game. In the homerun chase game, to swing the bat on game, a user swings an arm with an acceleration sensor. In the swimming game, when a user strokes with his arms with acceleration sensors, the game character swims forward. It took two hours for developing the homerun chase game and three hours for developing the swimming game. Most of the time for implementation was to create flash animations. Approximately 15 minutes are used for the definition of user motions and the description of rules that are needed to communications. Figure 7 shows snapshots of these games.

In addition, we created a content for donation box at Kobe Luminarie that is a festival held in Kobe, JAPAN. The festival is held to mourn the victims. We developed the content on the display in front of the donation box. The dog character in the content performs the same motion as a person who fold the donation box. For example, when the person bows, the dog on the display also bows and says, "Thank you for your donation" as shown in Figure 8.

This contents was also developed in approximately three hours. From these implementations, there is little time to implement interactive contents using sensors by our framework.

(a) waiting (b) bows

Fig. 8. Snapshots of the content in donation box

7 Conclusions

In this research, we proposed and implemented a framework that enables contents creators to create interactive contents using wearable sensors. The framework presents the mechanism for communication between Wearable toolkit and flash contents. Moreover, our proposed Code Generation Tool achieved automatic scripts/rules generation to use our framework easily. We confirmed that contents creator could develop intuitive contents in a few hours by evaluating actual implementations.

On the other hand, we feel that the difficulty of games becomes higher since the gesture input is difficult and delayed compared with the conventional input methods. Thus, we should adjust the game difficulty according to the recognition accuracy and the delay of the motion.

In the future, we will evaluate our framework based on actual implementations by flash content creators to show the effectiveness on complicated interactive contents.

References

1. Glovepie, http://carl.kenner.googlepages.com/
2. Makingthings, http://www.makingthings.com/
3. Phidgets, http://www.phidgets.com/
4. Adobe: Actionscript 2.0 components lauguage reference, http://livedocs.adobe.com/flash/9.0/main/
5. Adobe: Developer connection, http://www.adobe.com/devnet/flash/articles/
6. Adobe: Flash CS4, http://www.adobe.com/products/flash/
7. Dey, A.K., Hamid, R., Beckmann, C., Li, I., Hsu, D.: A Cappella: Programming by demonstration of context-aware applications. In: Proc. of International Conference on Human Factors in Computing Systems (CHI 2004), pp. 33–40 (2004)
8. Kobayashi, S., Endo, T., Harada, K., Oishi, S.: Gainer: A reconfigurable I/O module and software libraries for education. In: Proc. of the 2006 International Conference on New Interfaces for Musical Expression (NIME 2006), pp. 346–351 (2006)
9. Westeyn, T., Brashear, H., Atrash, A., Starner, T.: Georgia Tech Gesture toolkit: Supporting Experiments in Gesture Recognition. In: Proc. of International Conference on Perceptive and Multimodal User Interfaces (ICMI 2003), pp. 85–92 (2003)
10. MacIntyre, B., Gandy, M., Dow, S., David Bolte, J.: DART: A toolkit for Rapid Design Exploration of Augmented Reality Experiences. In: Proc. of Conference on User Interface Software and Technology (UIST 2004), pp. 197–206 (2004)
11. Kato, H., Billinghurst, M., Poupyrev, I., Imamoto, K., Tachibana, K.: Virtual Object Manipulation on a Table-Top AR Environment. In: Proc. of the International Symposium on Augmented Reality (ISAR 2000), pp. 111–119 (2000)

Research on Eclipse Based Media Art Authoring Tool for the Media Artist

Songlin Piao, Jae-Ho Kwak, and Whoi-Yul Kim

Image Engineering Lab
Division of Electrical and Computer Engineering
Hanyang University
Haengdang-Dong, Seongdong-Gu, Seoul, Republic of Korea, 133-791
{slpiao,jhkwak}@vision.hanyang.ac.kr,
wykim@hanyang.ac.kr
http://vision.hanyang.ac.kr

Abstract. A media art contents authoring tool based on Eclipse called Exhibition Contents Authoring System (ECAS) is presented in this paper. Visual editor of ECAS is implemented using Graphical Modeling Framework which is composed of Eclipse Modeling Framework and Graphical Editing Framework. The rest of the system were implemented using Eclipse Rich Client Platform framework. With this tool, artists could present their works easily by drag-and-dropping icons without programming skills.

Keywords: Media Art, Authoring, GMF, Eclipse, RCP.

1 Introduction

It is becoming easier and faster for the artists to create their works thanks to the development of computer technology. But there are still some limitations. One of them is that they still have to write programming code to some degree. However, most of professional artists do not know how to program even a single line of code. Our software Exhibition Contents Authoring System (ECAS), a media art contents authoring tool, has been developed for those artists in mind so that they could create their media art contents very easily. All they need to do is just drop and drag components of their choice, and then just connect these components in particular sequence by drawing lines between them. ECAS was developed using graphical modeling framework [1].

1.1 Related Work

There are several media art contents authoring tools available on the market in the form of commercial software or open source software. Max/MSP [2] and *Processing* [3] are two of the most widely used programs all over the world. Max

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 342–349, 2010.

was originally written by Miller Puckette as the patch editor for the Macintosh [4]. Then it was further developed by third parties to extend its functions gradually. There is an open source version of Max named Pure Data [5]. Pure Data, also developed by Miller, is an alternative tool for the students who learn digital music. *Processing* was firstly developed by the MIT Media lab in order to help out the artists who suffered from programming code. The development began in 2001 and the stable version 1.0 was firstly released by the year 2008. There are many media art contents developed and implemented using *Processing*.

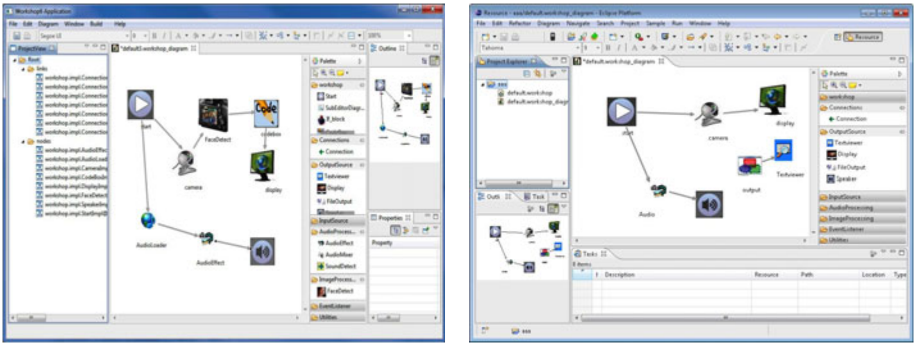
Besides these programs, there are also other media art contents authoring tools like VVVV [6], Quartz Composer [7] and Open Frameworks [8]. VVVV was designed to facilitate the handling of large media environments with physical interfaces, real-time motion graphics, audio and video that can interact with many users simultaneously [6]. Quartz Composer is a node-based visual programming language provided as part of Xcode in Mac OS X [4]. Open Frameworks is a C++ library designed to assist the creative process by providing a simple and intuitive framework for experimentation [8].

1.2 Motivation

Although the above tools are already convenient to use, there are still some obstacles that prevent media artists from easy creating their media contents using the tools. Max provides plenty of components, however, the structure is too complex and difficult for the artists to learn, not to mention the high price of commercial version. On the other hand, *Processing* is easier to learn and is free of charge, but instead of convenient graphical user interface, it still requires some programming code. VVVV has perfect rendering effect but it runs only on Windows platform.

There is a need to develop a cheap, easy to use, stable and efficient multi-platform media art authoring tool. Java and Eclipse were chosen to develop ECAS in order to get ability of multi-platform and higher extensibility. It is desirable that ECAS should not run only on the desktop PC but also run on the portable devices. It should also be easy to develop and maintain. As one of the world's most popular Integrated Development Environments, Eclipse has already provided these mechanisms. ECAS could be used as an independent desktop application or one of plug-ins [9] of Eclipse as in Fig. 1.

Although the code has been written in Java, there are several reasons that the speed of the program is as fast as in C/C++. Whatever Rich Client Platform (RCP) [10] or Graphical Modeling Framework (GMF), all the GUI frameworks from Eclipse are based on the SWT [11], which draws UI by calling system functions directly, therefore the speed grows up much faster than pure Java implementation. Just-in-time [12] compiling technology is used to compile the frequently used source code directly into the native code. While doing image processing and graphics rendering inside the software, JavaCV [13] and JOGL [14] are used. Both of these two libraries are implemented by calling native methods directly.



(a) ECAS as RCP (b) ECAS as Plugin

Fig. 1. Program appearance

The remainder of this paper is organized as follows. Section 2 discusses the whole architecture of proposed media art contents authoring tool. Subsection 2.1 introduces the procedure of designing model files using GMF. Subsection 2.2 shows the complementation procedure from two perspectives, one part is GUI and the other part is logic. The experimental result to evaluate the proposed media art authoring tool is described in Section 3. Section 4 gives the conclusion and future work.

2 Architecture

The system consists of two major parts: graphical user interface and program's logic. Usually, Graphical Editing Framework (GEF) [15] and Eclipse Modeling Framework (EMF) [16,17] are used for designing GUI and program's logic, respectively. But Eclipse provides more convenient way to generate GUI and program's logic simultaneously. The framework used here is GMF which is the combination of GEF and EMF. That is, instead of designing each part separately, infrastructure of the whole program is firstly designed using GMF and then the generated source code would be further modified manually to be fully functional.

2.1 Designing Using GMF

Domain model, graphical definition model, tooling definition model and diagram mapping model should be defined in order to develop GMF based infrastructure. The whole procedure involves following steps:

- Create a domain model, which defines the non-graphical information used and managed by ECAS
- Create a graphical definition model, which defines graphical elements to be displayed in the visual editor in ECAS

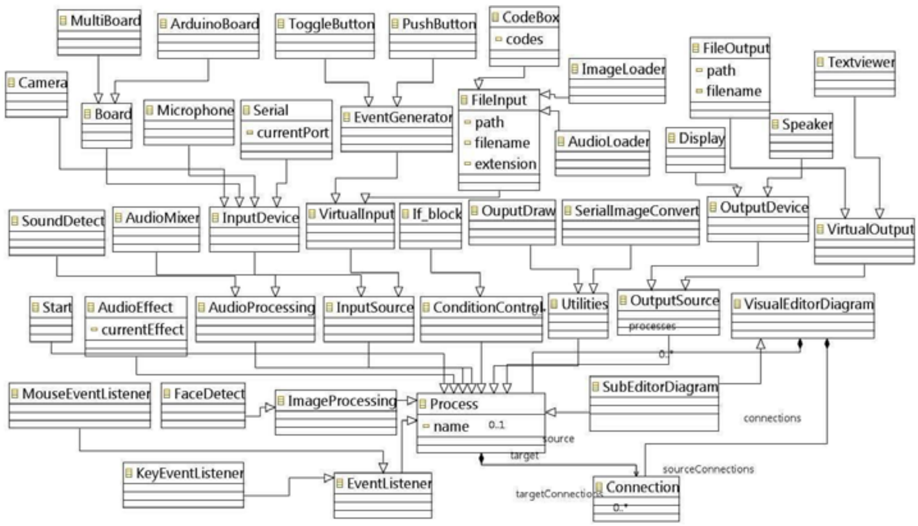


Fig. 2. ECAS function modeling

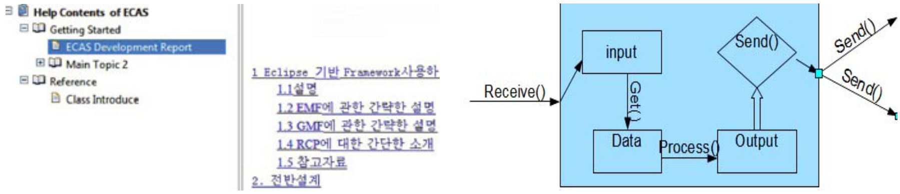
- Create a tooling definition model, which defines elements to be displayed in the palette in ECAS
- Create a diagram mapping model, which defines the mapping between domain model elements and graphical elements
- Generate the skeleton of graphical editor and enhance the generated code

First step is to create domain model, actually this is the part which is very closely related to program's logic. In the current version of ECAS, 42 classes are defined inside the domain model file (Fig. 2). The basic class here is named *Process*, because each block can be seen as a process when program is running. Seven methods are defined inside the *Process* class. They are listed below:

- void connect(*Process* target)
- void receiveMessage(Data data)
- void init() throws Exception
- void start()
- void process(Data param)
- void destroy()
- void stop() throws InterruptedException

After this domain model definition is done, a file which has the extension of *genmodel* is generated and then logic related source code is generated. The two main packages generated are workshop and workshop.impl. The definitions of the models' interfaces are located inside package workshop and the real implementation of the functionality is located inside the package workshop.impl.

Then the tooling definition model and graphical definition model should be defined. While designing tooling definition model, any number of groups could



(a) Help system (b) Process flow

Fig. 3. Help system and Process flow

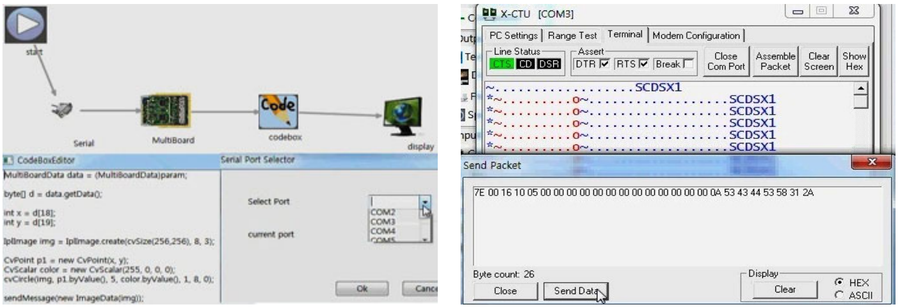
be created and each item is put inside the corresponding group. Icon image could also be changed by specifying the path of the image. In the case of graphical definition model, each node's figure should be specified inside the Figure Gallery. The default node shape is rectangle, but it could be modified to customized shape as in Fig. 1. Besides these, there are many functional options which could be implemented further. For example, whether put label inside or outside the node, whether fix the size of node or not and so on. The next step is to define diagram mapping model of these three files described above. After all these operations are finished, RCP based infrastructure was generated for the further development.

2.2 Complementation

The default generated source code is just the skeleton of the whole program. It already can run in this state, but the program cannot do real functions what artists want. Additional functions should be extended manually. As it was mentioned before, there are two main parts: one is GUI and the other is logic.

First part of program is about extensions on GUI. It is known that Eclipse provides many extension points. There are three main extensions implemented inside ECAS. The first one is *Action Set*, the second one is *View* extension and the third one is *Help* extension. Menu *Build* is added to the program by extending the *Action Sets*. There are two other submenus named as *Preview* and *Validation*. *Preview* is for running the program before doing export and *Validation* is for checking the integrity of the source file. These action sets are implemented as the subclasses of IWorkbenchActionDelegate class. We override the *run*() method of parent class. And the other extension is to extend the *View* in order to implement the project explorer. Project explorer can be seen on the left side in Fig. 1(a). Although the method used to implement project explorer is a little simple and itself can only provide limited functions, it do really very much help. In the future version, this scheme would be re-implemented using Common Navigator Framework (CNF). Help system is very important for any software. Help system inside ECAS is implemented from two side perspectives: manuals for the users and class references for the developers as in Fig. 3(a).

Second part of program is about extensions on program's logic. As it was mentioned earlier, the basic class of this Media art contents authoring tool is *Process*. It is defined as a thread. As every class used by ECAS is implemented



(a) (b)

Fig. 4. Multi-sensor board function test

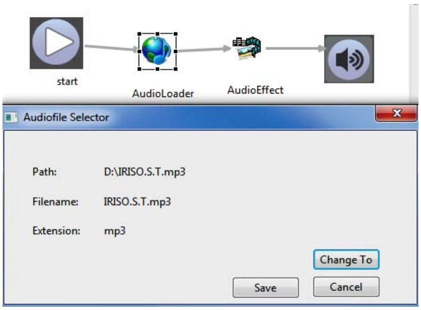
by inheriting thread class, it has its own run method. Fig. 3(b) shows the main concept of each *Process* class. The big box represents one node used inside the proposed tool. When the program is running, each node receives message from the previous node and stores the data to the *input* variable. Here *input* is a vector. In order to further process, the node first calls *get*() method to get the data from the *input* then uses *process*() method to do real logical processing, and then stores the result to the *output* variable. Finally, the node will call *send*() method to transfer the result to the next nodes. For example, if the node is Camera block, it receives the data from the local webcam, does some processing, and sends the result to the next node. When several nodes work together, we should provide the synchronization for these nodes.

3 Experiments

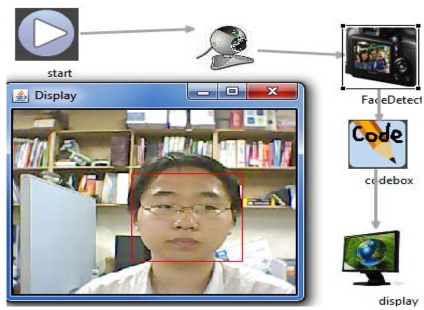
We did experiments on three aspects: external board, audio processing and video processing.

Fig. 4 shows the case for testing communication function block with external hardware, like Arduino Board [18] which is commonly used among the artists. The half above image in Fig. 4(a) shows the designed patch file and there are five blocks: start, serial, multi-board, code box, display. Start block means the start of the whole patch file; serial block is used for selecting the port which ECAS would use to communicate with multi-board; multi-board represents for the external device ECAS would communicate with; code box is designed for storing user defined code in order to deal with some problems more flexibly, in this case it will control how to display final result to the screen; display block is used for showing the result. The half nether image in Fig. 4(a) shows that artists could set or change some properties of each block through the prompt dialog box. After all the preliminary work is done, then we make the patch file running and Fig. 4(b) is the final result. Users can send and receive data through the interface presented in 4(b).

Fig. 5(a) shows the case for testing audio processing function block. There are four blocks in this case; start block, audio file block, audio effects block



(a) Audio test



(b) Video test

(c) Face detection in *Processing*

Fig. 5. Audio and Video processing function test

and speaker block. Start block is the same as in the previous case; audio file block is used for selecting the audio file for ECAS; audio effects block provides several kinds of effects, in this example only two are in use: *mixing* and *changing sampling rate*; speaker block is used for playing sound.

Fig. 5(b) shows the case for testing video processing function block. The test functionality in this case is face detection skill which is very commonly used in the image processing and computer vision area. The face detect function block is used for detecting faces inside one image frame. First, sequential images are grabbed from camera block then the camera block sends the data to the face detect block. The face detect block would detect human faces at each frame then send the result to the code box. Test code is set inside the code box block in order to retrieve the face information and draw it to the image buffer. Fig. 5(c) shows the same function implemented in *Processing*. Users need to write down about 50 lines code, but ECAS only uses five blocks.

4 Conclusion and Future Work

A media art authoring tool based on Eclipse is presented in this paper. The proposed scheme solves several problems of the existing tools. It can provide media artists more efficient and easier way of creating media so that they could create art contents very quickly and easily. Furthermore, it also makes the procedure easier for the developers so that they can improve this tool or maintain the existing code more efficiently.

Every block used inside ECAS is implemented as a thread. Although the thread can go to sleep when there is nothing to do, it still remains inefficient compared to those using static compiling method. Currently our users can only edit and run the program. But if there is an error inside, one cannot monitor exact situation. The function that users can also debug inside ECAS should be provided in future. The program is currently developed using GMF based on the Draw2D package. Although 2D is already enough for the artists to work, the mechanism that they work on 3D based graphical editor [19] would be added in the future, too. It is much more intuitive and user-friendly.

Acknowledgement. This research is supported by Ministry of Culture, Sports and Tourism (MICST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2009(2010).

References

1. Graphical modeling framework, http://www.eclipse.org/modeling/gmf
2. Cycling74 - tools for media, http://www.cycling74.com
3. Processing - open source programming language for media art, http://www.processing.org
4. Wikipedia, http://en.wikipedia.org
5. Puredata community site, http://puredata.info
6. A multipurpose toolkit, http://www.vvvv.org
7. working with quartz composer, http://developer.apple.com
8. open source c++ toolkit for creative coding, http://www.openframeworks.cc
9. Clayberg, E., Rubel, D.: Eclipse Plug-ins. Addison-Wesley Professional, Reading (2008)
10. McAffer, J., Lemieux, J.-M.: Eclipse Rich Client Platform: Designing, Coding, and Packaging Java(TM) Applications. Addison-Wesley Professional, Reading (2005)
11. Scarpino, M., Holder, S., Ng, S., Mihalkovic, L.: SWT/JFace in Action: GUI Design with Eclipse 3.0 (In Action series). Manning Publications Co., Greenwich (2004)
12. Suganuma, T., Ogasawara, T., Takeuchi, M., Yasue, T., Kawahito, M., Ishizaki, K., Komatsu, H., Nakatani, T.: Overview of the ibm java just-in-time compiler. IBM Syst. J. 39(1), 175–193 (2000)
13. Java interface to opencv, http://code.google.com/p/javacv
14. Java binding for the opengl, http://kenai.com/projects/jogl/pages/Home
15. Eclipse graphical editing framework, http://www.eclipse.org/gef
16. Eclipse modeling framework, http://www.eclipse.org/modeling/emf
17. Steinberg, D., Budinsky, F., Paternostro, M., Merks, E.: EMF: Eclipse Modeling Framework 2.0. Addison-Wesley Professional, Reading (2009)
18. Arduio home page, http://www.arduino.cc
19. Duske, K., von Pilgrim, J.: GEF goes 3D. Teil 3: 3D-Fizierung von GMF-Editoren. Eclipse Magazin 2(10), 79–82 (2010)

BAAP: A Behavioral Animation Authoring Platform for Emotion Driven 3D Virtual Characters

Ling Li[1], Gengdai Liu[2], Mingmin Zhang[1,*], Zhigeng Pan[1], and Edwin Song[3]

[1] State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China
{liling,zmm,zgpan}@cad.zju.edu.cn
[2] College of Computer Science, Xidian University, Xi'an, China
gdliu@xidian.edu.cn
[3] Intel Asia-Pacific Research & Development Ltd. Shanghai, China
edwin.song@intel.com

Abstract. Emotion, as an important aspect of human intelligence, has been playing a significant role in virtual characters. We propose an improved three-level structure of affective model as "personality-emotion-mood" for intelligent and emotional virtual characters. We also present the emotion state space, as well as the emotion updating functions, to generate authentic and expressive emotions. In order to achieve the complexity and variety of behaviors, we bring forward a behavior organizing structure as the behavior tree, which defines four kinds of behavior organizations as well as the behavior tag and behavior message, to manage virtual characters' behaviors. At the end, we achieve an experimental platform BAAP, which prove our emotion model and behavior organizing structure to be effective and practical in generating intelligent and emotional behavioral animations.

Keywords: virtual character, affective computing, behavior tree, behavioral animation, authoring platform.

1 Introduction

In previous work [1], we proposed a computational emotion model which can be incorporated into physiological and social components of emotions. We improve it in this paper and present an experimental platform BAAP. We mainly focus on two subjects: emotion modeling and behavior organizing. Based on the OCEAN model [2] and the OCC model [3], we propose an improved three-level affective structure as "personality-emotion-mood". The emotion state space, as well as the emotion updating functions, is defined for generating authentic and expressive emotions.

To organize behaviors of virtual characters, we bring forward a structure as the behavior tree similar to Halo2 [4]. There are four kinds of organizations for executing behaviors: "*Sequence*", "*Selector*", "*Repeat*" and "*Parallel*". By designing behavior tags and behavior messages, we succeed to generate intelligent and emotional behavioral animations.

* Corresponding author.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 350–357, 2010.

2 Related Work

Picard restates the definition of affective computing [5]. Information of personality and emotion can influence virtual characters over probability of choosing behaviors. Egges [6] presents the personality-emotion influence matrix to explain how each personality factor influences each emotion factor. Chittaro [7] defines an OCEAN-based model where behavior sequences are chosen from an animation library based on personality. The EMA model [8] combined a plan-based appraisal model with a detailed problem-focused model. However, one limitation among the previous work is the lack of interactions between physiology, cognition and social interaction.

Inspired by the psychological emotion models, researchers recognize the utility of computational models. Loyall [9] described an innovative system for authoring expressive, autonomous interactive characters. However, it focuses on building specific and unique believable characters, where the goal is artistic abstraction of reality, not biologically plausible behavior. Egges describe a generic model [10] and generate emotional, communicative and multiple character animations [11]. Perlin [12] create the autonomous digital actor to act out scenes using procedural animation techniques. All these works seek to create believable, emotional or social characters.

Behavior modeling and organizing are significant aspects of generating intelligent and emotional behavioral animations. Diller [13] focuses on behavior generations relevant for training applications. In the CreatureSmarts Architecture [14], robust, reactive, adaptable, honest, expressive, sensible and scalable creatures are sought to be built. Kallmann [15] presents a feature modeling approach to define behavioral information, paying special attention to the capabilities of interaction. Isla [4] uses hierarchical state machines and DAG [16] in Halo 2, where the use of state machines is regarded as less feasible for the long term and even proving difficult in generation games. The major problems are the complexity of state machines grow quickly when the behavior of a character becomes slightly more sophisticated.

3 Overview

Figure 1 shows the architecture of BAAP. In the first instance, a dialog system is provided for users to customize the virtual environment, for example, importing personal data (i.e. personality and emotion, which are described in detail in section 4) of virtual characters, customizing behavior tree based on the motion databases, and so on. Personal data of virtual characters is dealt with by the character controller.

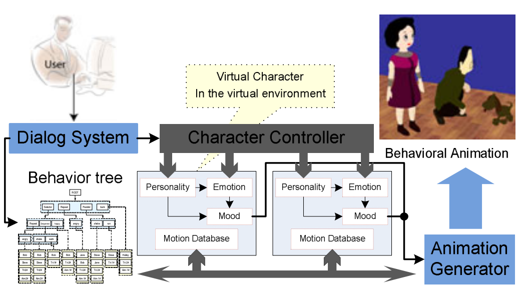


Fig. 1. Overview of BAAP

After initialization, the animation generator begins to take up the control. It first setup the virtual environment (i.e., locate virtual characters and objects, read in behavior data of the customized behavior tree and set initial emotion states and behaviors of virtual characters), afterwards, it carries out the execution of the behavior tree, which traverses leaf and non-leaf nodes of the behavior tree and executes behaviors recorded on the leaf nodes to generate behavioral animations. The particular introduction of behavior tree is described in section 5.

4 Emotion Modeling

4.1 Personality

Personality is closely relative with emotion and mood and is modeled using vector **p**, with each element corresponding to a component of the OCEAN model [2] and restricted within the interval **(0, 1)**:

$$\mathbf{p}^T = <\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3,\mathbf{a}_4,\mathbf{a}_5>, \ \forall \mathbf{i} \in \{1, 2, ..., 5\} \ \mathbf{a}_i \in (0, 1) \quad (1)$$

where $\mathbf{a}_i$ refers to each element of the OCEAN model.

4.2 Emotion

Emotion is formalized with an emotion state set of six elements (which are, *happiness*, *sadness*, *anger*, *fear*, *surprise* and *disgust* [17]) and specified using vector $\mathbf{e}_t$ at time **t** with each element restricted within the interval **(0, 1)**:

$$\mathbf{e}_t^T = <\mathbf{b}_1,\mathbf{b}_2,\mathbf{b}_3,\mathbf{b}_4,\mathbf{b}_5,\mathbf{b}_6>, \ \forall \mathbf{i} \in \{1, 2, ..., 6\} \ \mathbf{b}_i \in (0, 1), \ \mathbf{t}>0 \quad (2)$$

where $\mathbf{b}_i$ refers to each element of the emotion state set.

4.3 Emotion State Space

The emotion state space is designed in BAAP. As shown in Figure 2, six components are divided into three groups (namely, {*happiness*, *sadness*}, {*anger*, *fear*} and {*surprise*, *disgust*}), with the former of each represents the active component and the latter represents the passive component.

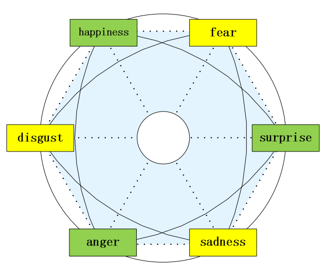


Fig. 2. Emotion state Space

The two components within the same group are independent of each other; however, they both hold either positive or negative relationships with components of other groups. Components of the same kind hold positive relationships with each other; while of different kinds hold negative relationships. For example, the active component "*happiness*" is independent of the passive component "*sadness*", holds positive relationships with the active components "*anger*" and "*surprise*", and negative relationships with the passive components "*fear*" and "*disgust*".

4.4 Emotion Updating

Different from the relative permanence of personality, emotion may change over time under rapid enhancing function $\mathbf{f_e}$ and slow weakening function $\mathbf{f_w}$:

$$\mathbf{f_e(b_i, u): b_i \leftarrow f(f^{-1}(b_i)+u), i \in \{1, 2, \ldots, 6\}, u>0} \quad (3)$$

$$\mathbf{f_w(b_i, v): b_i \leftarrow b_i*(1-e^{-1/vt}), i \in \{1, 2, \ldots, 6\}, v>0, t>0} \quad (4)$$

where $\mathbf{u}$ and $\mathbf{v}$ are the enhancing and weakening parameters respectively, $\mathbf{f^{-1}}$ is the inverse function of $\mathbf{f}$ described as follows:

$$\mathbf{f(x): x \leftarrow e^{-k_1/(x-k_2)}, k_1, k_2>0} \quad (5)$$

where $\mathbf{k_1}$ and $\mathbf{k_2}$ are parameters to determine the gradient of function $\mathbf{f}$.

4.5 Mood

Mood is virtual character's subjective experience. In this paper, only active mood and passive mood are considered. Mood is described with vector $\mathbf{m}$, with each component restricted within the interval $\mathbf{(0, 1)}$:

$$\mathbf{m^T = \langle m_a, m_p \rangle, m_a, m_p \in (0, 1)} \quad (6)$$

where

$$\mathbf{m_i = \log_3(\Sigma 3^{(bi-1)}), i\{a, p\}} \quad (7)$$

$\mathbf{m_a}$ and $\mathbf{m_p}$ denote the active and passive mood intensity respectively. $\mathbf{b_a}$ denotes the active components; while $\mathbf{b_p}$ denotes the passive components.

5 Behavior Organization

Behavior is the primary approach to represent emotion. The organization, formulation and execution of behavior are elaborated in this paper. The behavior DAG [16] (directed acyclic graph), namely, behavior tree, is implemented in BAAP. The non-leaf

Table 1. The organizations of behaviors in BAAP

organizations	descriptions
Sequence	All the direct children should execute in turn.
Selector	At most one direct child is chosen to execute.
Repeat	All the direct children should execute in turn for several times.
Parallel	All the direct children should start execution at the same time.

node of a behavior tree is designed to determine how to execute its subtrees; while the leaf node is the pure behavior to achieve some essential work. There are four kinds of non-leaf nodes, defining four kinds of organizations shown in Table 1.

5.1 Behavior Cinquain

There are five factors of behavior taken into account, namely *inflictor*, *sufferer*, *trigger*, *aim* and *state*, which compose behavior cinquain **bhv**:

$$\mathbf{bhv = <inf, suf, tri, aim, state>} \quad (8)$$

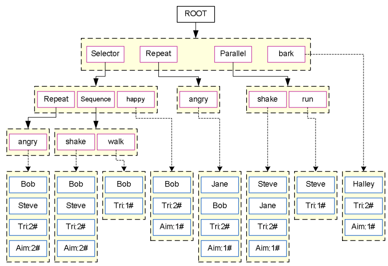


Fig. 3. Example Behavior tree

Figure 3 shows an example behavior tree, with one and only root node. The pink rectangles indicate leaf and non-leaf nodes; while the blue ones contain elements of pure behaviors. Moreover, the words recorded in each top blue rectangle (i.e. the leftmost "*Bob*") denote the inflictor; while the second ones (i.e. the leftmost "*Steve*") denote the sufferer; such words as "*Tri: 2#*" indicate that there are two components recorded in the **tri** element.

5.2 Behavior Tag

It is difficult to prepare everything in detail when customizing a behavior tree, for the complexity of solving behavior factors will grow implausible if the tree becomes massive. Luckily, we have the behavior tag. The idea is not to keep all these specific factors in the behavior cinquain but encode some of them as a tag, which are ignored when customizing the behavior tree. Not until the behavior is executing will the tag be computed, in this case, all the details of the behavior can be obtained.

5.3 Behavior Message

There is another problem, that is, once a behavior finishes execution, the situation of the virtual environment may have changed. Then, how to inform the changes?

Our method is through the behavior message, which does not appear in a static behavior tree structure, but is instead dynamically generated by the message handler to a specific behavior node during the behavior tree is executing. Concretely, a behavior calls the message handler after executing; the message handler, which conserves current environment information, picks up this call and releases a message containing the updated information; virtual characters will update their information accordingly and respectively once receiving the message.

6 Results and Discussions

Two contrastive experiments, where virtual characters are customized with different personalities and emotions, are prepared in BAAP. To obtain convincing results, we make them share the same behavior tree; unfortunately, we are only able to affix two small segments of the tree because of its large scale, as shown in Figure 4.

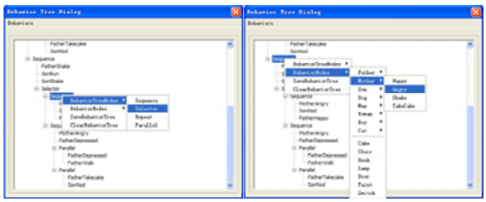


Fig. 4. Segments of the behavior tree built in BAAP

In Experiment 1, the personality of Bob is customized to be high in neuroticism, while Jane in openness and extraversion, Steve in agreeableness and Halley in conscientiousness. As for initial emotions, we customize Bob to be high in fear, surprise and disgust, Jane to be in anger, Steve to be in surprise and Halley to be neutral. The reason for Halley to be neutral is we intend to build a reference.

As shown in Figure 5, a segment of behavioral animations is generated through the execution of the behavior tree. The general process is: 1). Halley keeps on barking at the cake, which attracts Steve. 2). Steve fails to reach the cake and runs to Bob. 3). Steve asks Bob for help. 4). Bob refuses Steve's request. 5). Steve does not give up and runs to Jane. 6). Steve asks Jane for help. 7). Jane gets angry with Bob. 8). Bob feels dejected and walks to the cake. 9). Steve feels pleased for Bob's help.

Fig. 5. Procedure of Experiment 1

In Experiment 2, the personality of Bob is customized to be high in openness and agreeableness, while Jane in extraversion, Steve in agreeableness and Halley in conscientiousness. In contrast with Experiment 1, we customize Bob to be high in happiness, Jane to be in anger, Steve to be in fear and Halley still to be neutral.

As shown in Figure 6, a contrastive segment of behavioral animations is generated through the execution of the behavior tree. The general process is: 1). Halley keeps on barking at the cake, which attracts Steve. 2). Steve fails to reach the cake and runs to Bob for help. 3). Bob approves Steve's request. 4). Bob feels happy and walks to the cake; Steve follows Bob with pleasure and excitement. 5). Bob takes the cake. 6). Steve feels satisfied with Bob's help.

Fig. 6. Procedure of Experiment 2

We achieve two contrastive segments of behavioral animations. With neurotic personality and disgusted emotion, Bob will refuse Steve's request (in Experiment 1), however, it will approve Steve's request if customized to be open and agreeable (in Experiment 2). Similarly, Jane and Steve will also behave differently with different personalities and emotions. Driven by personalities and emotions, our method enables virtual characters to present authentic and expressive behaviors.

Figure 7 shows the mood curves of virtual characters in Experiment 1 (the left picture) and Experiment 2 (the right picture). In the left picture, we are informed that Bob goes through an emotional process with twists and turns (the red curve); Jane keeps a mood curve with a sudden rise (the blue curve); Steve's mood curve arises right after Bob's first rise (the green curve). We get the opposite result in the right picture where Bob goes through a flat emotional process (the red curve) and Steve looks with twists and turns (the green curve).

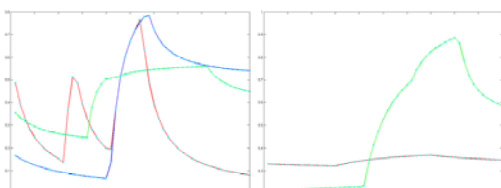


Fig. 7. Mood curves of virtual characters in the two contrastive experiments

7 Conclusions and Future Work

In this paper, an improved three-level structure of affective model as "personality-emotion-mood", as well as a new behavior organizing structure behavior tree, is brought forward to generate intelligent and emotional behavioral. An interactive platform named BAAP is also constructed to allow users to customize virtual characters to achieve authentic and expressive emotions as well as behaviors. Our future work will focus on facial expressions and learnable abilities of virtual characters. In addition, we believe that better behavioral animations will be generated with better technology of motion controllers and physical simulations.

Acknowledgements

This research is co-funded by INTEL-University cooperation project and the 863 project (Grant NO: 2006AA01Z303).

References

1. Yang, H., Pan, Z., Zhang, M.: Modeling emotional action for social characters. Knowledge Engineering Review 23(4), 321–337 (2008)
2. Larstone, R., Jang, K., Livesley, W., Vernon, P.: The relationship between Eysenck's P-E-N model of personality, the five-factor model of personality, and traits delineating personality dysfunction. Personality and Individual Differences 33, 25–37 (2002)
3. Ortony, A., Clore, G., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge (1988)
4. Isla, D.: Handling Complexity in the Halo 2 AI. In: Proceedings of Game Developers Conference, GDC (2005)
5. Picard, R.: Affective computing: challenges. International Journal of Human Computer Studies 59(1-2), 55–64 (2003)
6. Egges, A., Kshirsagar, S., Thalmann, N.: Generic personality and emotion simulation for conversational agents. CAVW 15(1), 1–13 (2004)
7. Chittaro, L., Serra, M.: Behavioral programming of autonomous characters based on probabilistic automata and personality. CAVW 15(34), 319–326 (2004)
8. Marsella, S., Gratch, J.: EMA: A computational model of appraisal dynamics. In: European Meeting on Cybernetics and Systems Research (2006)
9. Loyall, A., Reilly, W., Bates, J., Weyhrauch, P.: System for authoring highly interactive, personality-rich interactive characters. In: Proceedings of the 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 59–68 (2004)
10. Egges, A., Kshirsagar, S., Thalmann, N.: A Model for Personality and Emotion Simulation. In: Palade, V., Howlett, R.J., Jain, L. (eds.) KES 2003. LNCS (LNAI), vol. 2774. Springer, Heidelberg (2003)
11. Egges, A., Thalmann, N.: Emotional communicative body animation for multiple characters. In: Proceedings of the First International Workshop on Crowd Simulation, pp. 31–40 (2005)
12. Perlin, K., Seidman, G.: Autonomous Digital Actors, Motion in Games (2008)
13. Diller, D., Ferguson, W., Leung, W., Benyo, A., Foley, D.: Behavior modeling in commercial games. In: BRIMS 2004: Proceedings of the 2004 Behavior Representation in Modeling and Simulation Conference (2004)
14. Burke, R., Isla, D., Downie, M., Ivanov, Y., Blumberg, B.: CreatureSmarts: The Art and Architecture of a Virtual Brain. In: Proceedings of the Game Developers Conference (2001)
15. Greisemer, J., Butcher, C.: The Integration of AI and Level Design in Halo. In: Proceedings of the Game Developers Conference, San Jose, CA (2002)
16. Kallmann, M., Thalmann, D.: Modeling Behaviors of Interactive Objects for Real-Time Virtual Environments. Visual Languages and Computing 13, 177–195 (2002)
17. Ekman, P.: An argument for basic emotions. Cognition and Emotion (1992)

Choshi Design System from 2D Images

Natsuki Takayama, Shubing Meng, and Takahashi Hiroki

Graduate School of Informatics and Engineering,
University of Electro-Communications,
1-5-1 Chofugaoka, Chofu-shi, Tokyo
msb8063@yahoo.co.jp, rocky@hc.uec.ac.jp

Abstract. This paper proposes a Choshi design system. Choshi is a new method for carving of paper, which is uneven 3D shape and unique colors of papers. Choshi, derived from carving overlaid colored papers, has the following three features:

1. Each layer consists of a single piece of paper of one color.
2. The color must be selected from a number of existing colors.
3. Choshi has an overlaid structure where carved papers are overlaid on other carved papers.

The goal of the proposed Choshi design system has two issues: to enable a wider variety of people to easily and successfully create Choshi art, and to reduce the difficulty and tedium, of creating a Choshi art piece.

Keywords: Choshi, Segmentation, Selecting colored paper, Overlaid structure.

1 Introduction

Choshi[6] is a new carving method which can represent uneven 3D structures with unique colored papers as shown in Figure 1.

The process of Choshi art, shown in Figure 2, includes:

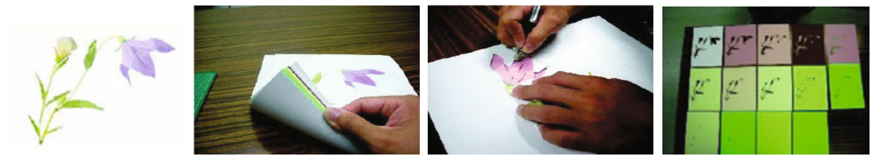
1. A color sketch should be prepared in advance. A number to each colored regions corresponding to the carving sequence is assigned.
2. Papers which are similar colors to those in the sketch are selected. All colored papers are put together according to the order.
3. The topmost layer should carve away the color regions of all the other colors. The next layer should carve away the color regions of all the remaining colors. This proceeds layer by layer to the bottom. The visible color regions in all papers result in a color sketch with sharp curves and a complex overlapping structure.

Compared with graphic works and three-dimensional arts, Choshi provides a better aesthetic perception to the viewers. It is, however, very difficult to popularize this art form due to its sophisticated structure, complicated and time-consuming production processes, and high requirement for the artistic knowledge.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 358–365, 2010.

(a) Choshi top view image (b) Choshi side view showing layers image

Fig. 1. Choshi



(a) A sketch (b) Colored papers (c) Regions carving (d) Discrete papers

Fig. 2. Making processing of Choshi

NPR is the short form of Non-Photorealistic Rendering. It assists artists in creating real artworks starting with digital images, and provides the possibility free the artists from labor of complex handwork and get the same or even better artistic effect than that made by hand. Some of previous NPR methods which are used for representing three dimensional shapes, similar to Choshi, are introduced.

Bas-relief is a form of relief in which the contents silhouette slightly stands out from the surrounding surface. The overall depth of bas-relief is shallow so it sets up an intermediate step between 2D image and 3D sculpture. For example on most coins, images are in bas-relief. A presented digital bas-relief system[4] assists artists in converting virtual 3D models into digital bas-relief sculptures. During the conversion the input 3D models should be squeezed into a flat surface, while maintaining as much as possible the perception of the full 3D scene. As an NPR system it can generate real bas-relief automatically with the aid of computer-driven milling equipment. In addition, this bas-relief system has a few functions to assist artists in adjusting aesthetic qualities such as the camera viewpoint or the compressed height in order to create digital Bas-relief intuitively based on the their personal aesthetic perception.

Another virtual sculpture system[3] allows artists to realize traditional sculpture in virtual space by CSG(Constructive Solid Geometry) and height-map. The system provides an intuitive interactive interface to artist. Artists can carve any complicated polygon by using a digital graver to cut forms on stone, wood, brick or any other material in virtual space.

Neither digital Bas-relief nor virtual sculpture can be used in Choshi generation for the discrete overlapping structure of color papers.

The rest of paper is organized as follows; section 2 describes Choshi system by giving the details of each process, and describes the application of the Choshi design system to the generation of carving data. Also, some real Choshi are carved by a cutting plotter based on the data. Section 3 concludes this paper.

2 Choshi Design System

2.1 Generation of Rough Sketch

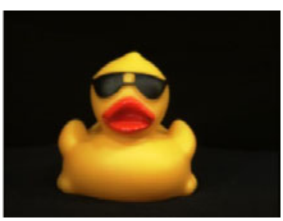
Region segmentation of Mean Shift algorithm[7] can not only keep the details and structures of the original features in images but also can facilitate the labeling process better. In this paper, $q = (p, c) = (x, y, r, g, b)$ is used as a feature vector, where $p = (x, y)$ represents a vector in spatial domain H^2, and $c = (r, g, b)$ represents a vector in color space S^3. In the five-dimensional feature space, mean shift vector $M(q)$ is obtained by the following equation.

$$\boldsymbol{M(q)} = \frac{\mathbf{1}}{\boldsymbol{n_x}} \sum_{\boldsymbol{p \in H^2, c \in S^3}} \boldsymbol{(q_i - q)} \tag{1}$$

Where, q and q_i denote five-dimensional feature vectors of the data. n_x represents the number of data within the radius of $d_H \in H^2$, $d_S \in S^3$ at the center q.

Local mode is obtained by shifting the target vector $\boldsymbol{q}$ until the mean shift vector $M(q)$ becomes smaller than a threshold. This process is applied for the whole pixels in a given image. Color value in each pixel is replaced the mode in the local distribution. The adjacent pixels in spatial domain are merged into one region, where the distances in the feature space are within the range of d_H and $\frac{d_S}{\sqrt{2}}$.

The adjacency relation of regions in the image space is represented by RAG. Color difference within the range of $\frac{d_S}{\sqrt{2}}$ in the color feature space is merged into one region. This process is repeated 3 times. The obtained image is used as rough sketch as shown in Figure 3(a).


(a) Input image

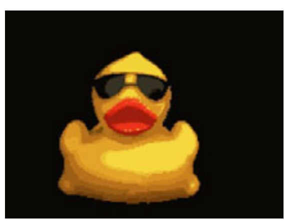

(b) Segmented image

Fig. 3. Result of rough sketch image

2.2 Colored Paper Selection

In representation of the painting style like Choshi, it is most desirable to calculate color difference in a color coordinate system which is close to human color perception. Munsell color system has three axes H(Hue), V(Value) and C(Chroma) which indicate three attributes of human color perception respectively.

The representative color value of each color region in the rough sketch and colored papers are represented by HVC coordinate in approximate Munsell color system by using MTM[9]. Then the color differences between the representative color in the rough sketch and the colored paper are calculated by Godlove Color different formula. A colored paper is selected when the color difference is minimum.

MTM corrects the evenness of Adams color space which approximates Munsell color space. At first, RGB color represented 8 bits in each component is transformed into CIE(1934)XYZ color value by using the following formula.

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \frac{1}{255} \begin{pmatrix} 0.608 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.144 \\ 0.0 & 0.066 & 1.112 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{2}$$

Next, Munsell Value function $V(A) = 11.6A^{\frac{1}{3}} - 1.6$ is applyed to calculate (M_1, M_2, M_3) as three components in Adams color space according to equation (3).

$$\begin{cases} M_1 = V(1.02X) - V(Y) \\ M_2 = 0.4[V(0.847Z) - V(Y)] \\ M_3 = 0.23V(Y) \end{cases} \tag{3}$$

M_1 and M_2 indicate color plane and M_3 is for Value. The approximate value $(\hat{H}, \hat{V}, \hat{C})$ of HVC in Adams color space is given by the following equation(4).

$$\hat{H} = tan^{-1}\left(\frac{M_2}{M_1}\right), \quad \hat{V} = M_3, \quad \hat{C} = ({M_1}^2 + {M_2}^2)^{\frac{1}{2}} \tag{4}$$

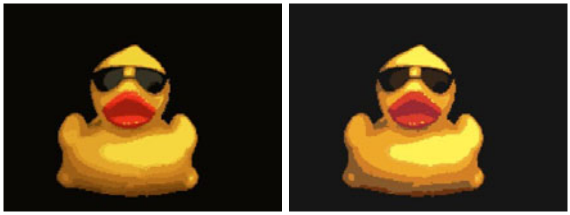
Then, new axes S_1 and S_2 are obtained by equation (5) which aims to correct the distortion of M_1 and M_2.

$$\begin{cases} S_1 = 8.88 + 0.966cos(\hat{H})M_1 \\ S_2 = 8.025 + 2.558sin(\hat{H})M_2 \end{cases} \tag{5}$$

Approximate value of HVC derived from MTM can be given by the following equation by using S_1, S_2 and M_3.

$$\tilde{H} = tan^{-1}(\frac{S_2}{S_1}), \quad \tilde{V} = M_3, \quad \tilde{C} = ({S_1}^2 + {S_2}^2)^{\frac{1}{2}} \tag{6}$$

Godlove color difference formula gives weighted Euclidean distance between the two color samples in Munsel color space, where the distance is weighted by the perceptual ratio of C and V. Godlove color difference formula is shown as



(a) Segmented image (b) Final reference image

Fig. 4. Result of colored paper determination

equation(7), where a perceptual ratio of C and V is $C : V = 1 : n$. In this paper, n is set to 4 according to the experimental result of Godlove[10].

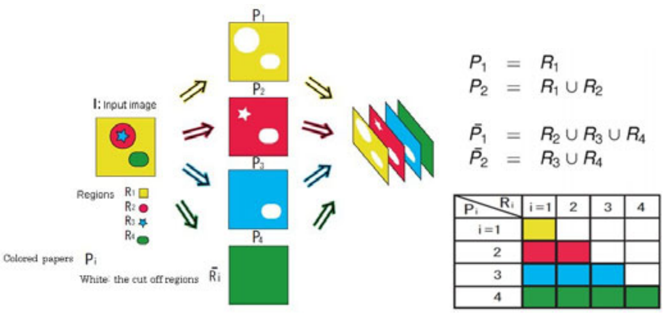
$$\Delta E = ((\Delta \tilde{C})^2 + 2\tilde{C}_1\tilde{C}_2(1 - cos\Delta \tilde{H}) + (n\Delta \tilde{V})^2)^{\frac{1}{2}} \quad (7)$$

Figure 4 shows the selected colored paper according to equation (7).

2.3 Choshi Description

In every colored paper, the regions which are not seen also exist, so the overlapping structure must be considered. Based on the result of color region segmentation, the whole image is indicated as I, and the set of all the color regions in the image are indicated as $R_i \in I(i = 1, 2, ..., n|n)$, n indicates how many pieces of paper in use. For $I = \bigcup_{i=1}^{n} R_i$, $\bar{P}_i$ indicates the cut-off part in each colored paper. P_i denotes the left part. When the papers are piled up in the order of i, P_i is represented as follows;

$$\begin{cases} P_1 = & R_1 \\ P_i = P_{i-1} \cup R_i & (i = 2, ..., n) \end{cases} \quad (8)$$



(a) Overlapping structure (b) Relationship between regions and colored papers

Fig. 5. Determine the sequence of colored papers

The cut-off part of each colored paper $\bar{P}_i$, therefore, derives from $\bar{P}_i = I - P_i = \bigcup_{k=i+1}^{n} R_k$ obtained by the following equation;

$$\begin{cases} \bar{P}_n = & \phi \\ \bar{P}_i = \bar{P}_{i+1} \cup R_{i+1} & (i = n-1, ..., 1) \end{cases} \quad (9)$$

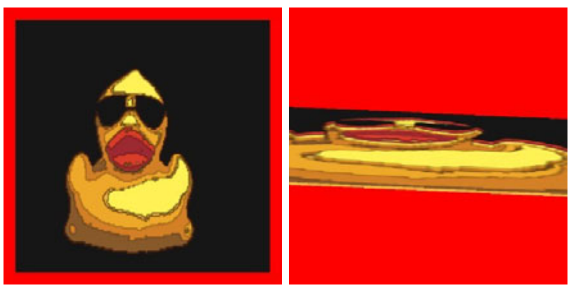
The overlapping structure model is illustrated in Figure 5.

The proposed Choshi system assigns certain numbers to the color regions based on the sequence which is determined by the order of generated regions in the rough sketch generation process. The number indicates a height.

After the selection of colored papers, user can adjust the height of each colored paper by exchanging the numbers of the regions. Moreover, user can merge some of the regions in order to modify the shapes of regions and the number of colors.

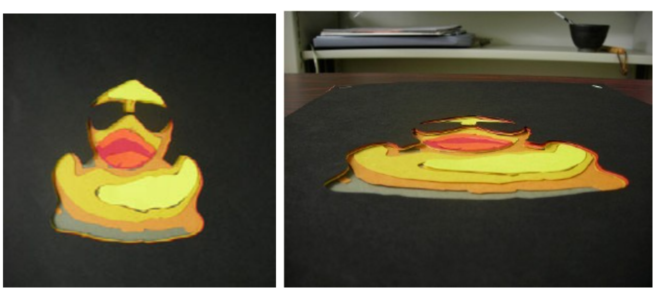
It is realized by interactive operations that are mouse operation and number assignment. The Choshi model is generated in a virtual space based on the equations (8) and (9) as shown in Figure 6.

Then, the cutting plotter is used to cut each colored paper according to the output data and the colored paper is piled up in the order. Finally, a real Choshi can be obtained. The generated Choshi is shown in Figure 7.



(a) Viewed from front (b) Viewed from bottom

Fig. 6. Virtual Choshi model



(a) Viewed from front (b) Viewed from bottom

Fig. 7. Carved Choshi

3 Conclusion and Future Works

The system proposed in this paper makes it possible to generate virtual Choshi beginning with an input image. It takes into account the following three features of Choshi in order to keep the generated overlapping structure as similar as possible to the image.

1. Choshi is composed by a set of colored papers and each color region is represented by a piece of colored paper. Each region, therefore, should hold a uniform color.
2. Segmented regions show many kinds of colors, but the resulting Choshi will be realized by colored papers. Thus the categories of colors have to be limited to the existing colored papers.
3. In every colored paper, the regions which are not seen also exist, so the overlapping structure must be considered.

Moveover, as an automatic method, it provides users a simple and pleasant way to make Choshi. It is helpful to spread the Choshi as a common art form.

However, this system is not mature enough. It still needs further improvement.

In proposed system, the quality of the result is very dependent on the quality of the segmentation. Experiment results show that the Mean Shift algorithm is more adapt for defining color regions of illustration as shown in Figure 8(a), however, it does not seem to robustly define ideal regions of input image with complex scenes, like photograph. As shown in Figure 8(b), the discrete leaves and windows did not get enough merging for matching unique color regions of Choshi.

(a) Choshi model issued from an illustration

(b) Choshi model issued from a photograph

Fig. 8. Experimental results

In addition, the Mean Shift algorithm can not intelligently judge the relationship among regions in an image. Some unnatural effects exist as the results of incorrect judgement of automatic segmentation, like the separated sky regions shown in Figure 8(b).

Such points will lead to the results that a Choshi artist would not be satisfied with. Therefore, it might be preferable to provide the artist with tools to guide the segmentation process and edit the results, rather than to rely on the described region segmentation algorithm.

References

1. Miyada, K.: Technical trend of rendering. The Journal of the Institute of Image Information and Television Engineers 57(4), 442–443 (2003)
2. Takagi, S.: Research trend of non-photorealistic rendering. Technical report of IEICE. HCS 23(63), 1–6 (1999)
3. Okada, M., Ichiro, Toriwaki, J., Mizuno, S.: Virtual sculpting and virtual woodblock printing by a model-driven scheme. Information Processing Society of Japan 1(2), 74–84 (2002)
4. Weyrich, T., Deng, J., Barnes, C., Rusinkiewicz, S., Finkelstein, A.: Digital bas-relief from 3D scenes. ACM Transactions on Graphics (Proc. SIGGRAPH) 26(3), 32–38 (2007)
5. Ikuta, D., Tanaka, T., Horiba, I., Yamamoto, S., Suqie, N.: Relief design system with multiple modeling methods. Technical Report of IEICE 99(79), 19–24 (2000)
6. Hayashi, K.: Choshi arts. Patent Laid-Open Number: 2004-255858 (2004)
7. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Trans. on PAMI 24(5), 603–619 (2002)
8. Tremeau, A., Colantoni, P.: Regions adjacency graph applied to color image segmentation. IEEE Trans. Image. Process 9(4), 735–744 (2000)
9. Miyahara, M., Yasuhiro, Y.: Mathematical transform of (r, g, b) color data to munsell (h, v, c) color data (high efficiency coding technic for recording and transmission). The Journal of the Institue of Television Engineers of Japan 43(10), 1129–1136 (1989)
10. Godlove, I.: Improved color-difference formula, with applications to the perceptibility and acceptability of fadings, pp. 760–772

Player's Model: Criteria for a Gameplay Profile Measure

Emmanuel Guardiola and Stephane Natkin

CEDRIC /CNAM, 292 Rue Saint-Martin
75003 Paris, France
emmanuel.guardiola@seasideagency.com, stephane.natkin@cnam.fr

Abstract. Game designers empirically use psychological and sociological player's model to create the gameplay of their video games. These models are generally implicit and always informal. A formal analysis of the player's model leads to define efficient player behavior profile. It can have numerous applications, for instance adaptation of the content to the player's ability and interest. Our work tries to find a rational way to assess Players Styles, concept suggested by Bartle [1] in 1996. The first step, state of the art of the player model, shows already some interesting criteria that can be used to classify player's styles.

Keywords: Video games, player's model, player's profile, gameplay, game design.

1 Introduction: The Playing Player

Bartle was the first designer trying to analyze the player's behavior from a gameplay point of view. His classification of players in MUD (Multi User Dungeon) is still a reference discussed both by designers and sociologist. Bartle's model is based on interviews done after game sessions. A methodological potential flaw is related to the fact that data were collected out of the magic circle [4]. As a consequence any measurement should investigate the specific playing time, when the player feels engaged [2] and perceives his posture as *an active player* [3]. So the measure has to be performed during active gameplay.

The main criticism regarding Bartle's model is related to contextual evaluation of the player's behavior. It can only reveal MUD players' profiles for a given game and at a given period. So the player model may be built on general criteria already used to create all sort of game and more generally based on general behavioral theories.

Our goal is to analyze criteria needed to assess the player while he's playing and reacting to gameplay mechanisms. Therefore we should rely on a definition of Gameplay.

Gameplay definition: Gameplay is often defined as the relation between player and game world as Salen and Zimmerman evoke it [5]. On one hand we have a designed game system that simulates something. On the other hand, we have a player, creating personal mental imagery of the situation and deciding mentally how to act within. Between the game system and the player, interfaces and captors create the link. In the

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 366–371, 2010.

gameplay situations, sometimes called patterns [6], the player has a goal, a kind of challenge to reach that goal while facing uncertainty about how the game will end [7].

Therefore, the definition of gameplay used in this research should be: All actions (cognitive or physical, performances and strategies) performed by the player, influencing negatively or positively the outcome of the uncertain game situation in which he is immersed in.

2 Player Model in Gameplay State of Art

Model of the playing player can be explored from several empirical and academic points of view. This paper focuses on the constructive aspects of the harvesting.

Empirical difficulty and Flow. Game developers spend important resources in design, test and tuning to insure that the player is always confronted to an interesting challenge, not boring and not too frustrating. All these expenses seem to be legitimized by Milahy Csikszentmihalyi theory of flow [8]. Difficulty test sessions include a model of the player built upon pragmatic aspects of his gameplay: how much time he takes to end a level, number of "game over" in a level… From data collected, developers design the path through their games for a generic player or several classes of players if the game offers several levels of difficulty. Some games, as Oblivion [9], implement a dynamic difficulty system linked to the player's performance in the game.

Two criteria help measuring the relationship between goal and effective result: Efficiency and Effectiveness. Efficiency is: how much resources are spent to reach the player's goal. For example, many "Survival horror" are fully based on the quality of resources management (rare ammunitions, inventory). Effectiveness is: does the player reach the goal (or how long did it take to reach the goal)? Another possible criterion to be measured in test lab is perseverance: How many times a player retries a given challenge? Does the player focus on one quest or micro challenge until he succeeds or does he try to find alternatives?

Empirical critical path and bonus. In almost all game levels or even challenges there are several ways to win. Some are based on the "longest path": In most solo action adventure games, first person shooter, role playing games, designers create a critical path: A walkthrough across all the levels that allows the player to meet with most of the assets produced. But, if the player rushes he cans win using a "shortest path". As a consequence he missed potential bonus, hidden rewards and secret levels. Some players want to end the game, others want to see, know or get everything. An evaluation of the ratio of the number of events, situations, challenges encountered by the player using his strategy and the "longest past" ones can be considered as characteristics of player's behavior. This is, in Bartle's classification difference between Killers and Explorers, for example.

Performance modeling. All games, according to Jesper Juul's definition, include a scoring system. A scoring system is generally a poor indicator of the player's behavior but it can be enhanced in several ways. In the game Forza Motorsport [10] the Drivatar feature is an efficient way to model the player's performance [11]. As he drives the player trains an AI. The system records a few key characteristics linked to gameplay as how consistent you are in, your speed and line or at what point you brake before entering a turn. Then you can let the AI drives a car with some of your gameplay attributes. This research shows spectacularly that cloning gameplay performance is possible with a very few variables to track.

Artificial intelligence. Artificial intelligence research gives several points of view on the player's model. The goal can be, for instance, to adapt challenge to the player or create more human behavior for non-player's character.

Historically in Chess and Go AI, the modeling process target is to create a virtual player who can find the optimal solution in a challenge. But real players are not always optimal in their choices.

Another approach is based on tracing the strategies used in specific contexts. It allows for creation of an Action Model with identification of strategy through gameplay patterns. All gameplay's variables must be taken in account. This is easy to do with games like Pacman [12] or Space invader [13] which rely only on a few features. It may be much more complex to realize on elaborates gameplay and in particular on sandbox games. The pattern recognition principle is mainly applicable to a priori defined strategies. How could a computer classify strategies that were not forecasted by the designer?

If strategies preferences are good criteria to model the player, we need a complementary way to capture player's style.

Decision theory. Decision Theory provides an interesting behavioral model of decision under risk, [14]. This scientific model matches with a classification of poker players used by players themselves to guess competitors' strategies: Risk-seeking for gains of low probability, Risk-aversion for losses of low probability, Risk-aversion for gains of high probability, Risk-seeking for losses of high probability. Customizing that approach for video game, we may classify this behavior in a two dimension scheme by measuring the percentage of success of challenges together with player involvement in it.

Two other notions linked to Decision Theory can be taken into account and added to our Gameplay criteria. The Utility principle: 5 $ do not have the same value for everybody. Players evaluate subjectively the value of the resources available and potential rewards. The ambiguity principle: we take decision with a certain amount of information on the situation, a number of known variables. What amount is acceptable for us?

Motivation Theory. Guillaume Denis [16] uses the motivation continuum of Deci and Ryan [17] in video game context. We can identify player progression from amotivation to intrinsic motivation. Role playing games as World of Warcraft (WOW) [18] provide several layer of motivational content to help the player progression through different levels of external regulations (See table 1). At the integrated regulation level, we can identify Bartle's player classification.

Game Theory. In a previous work, we used game theory to analyze gameplay decision [15]. A game situation can be sorted out in a finite number of matrices. We can track player choices and try to define the pay off. The conclusion of this experiment reveals a possible way to differentiate short, medium and long term behaviors.

Table 1. Deci and Ryan motivation continuum table with WOW regulation progression

Behavior	Non self determined	Continuum →				Self determined
Type of motivation	Amotivation	Extrinsic motivation				Intrinsic motivation
Type of regulation	Non regulation	External regulation	Introjected regulation	Indentified regulation	Integrated regulation	Intrinsic regulation
Locus of causality	Impersonal	External	Somewhat external	Somewhat internal	Internal	Internal
WOW progression of regulation		Gold Drop XP points…	Quest solving, Level up, Characteristic and competence progression…	Distance between action and goal : creating a wanted item= collect these resources + acquire this competence + buying this element	Explore everything, achieving all quests, being the best in PV, being pacifist defend guild principles, …	

3 Criteria of the Player Gameplay Profile

From the previous section, we can identify several interesting criteria to establish a player's gameplay profile.

Table 2. Criteria for a gameplay profile

Performance	In this particular gameplay, how the player performs?
Efficiency	How much resources I spend to reach the goal.
Effectiveness	Do I reach the goal?
Strategies preferences	What type of strategies the player knows, uses, masters. What context/strategies pattern he reproduces
Completion	What % of the reachable game he finished
Perseverance	What is his degree of retrying facing failure?
Risk aversion profile	What is the position of the player in the risk aversion area?
Utility	What is his level of utility, what are the valuables elements for him?
Ambiguity	Does he take risk in uncertainty or knowing all the variables?
Motivation type	Where is the player on the motivation continuum?
Pay off term preference	Is he short-term, medium-term or long term in his choices?

The most pragmatic trace of player's decision is the input. This information must be interpreted through the filter of the game system to understand its gameplay meaning. We work with the company Mimesis Republic on their virtual playground Mamba Nation [19] to confront this list to real data records. To be efficient, criteria were transformed in dimensions linked to clear player input. This work brings out several remarks.

Performance measure must be elaborated for each game, and in the game, for each gameplay. Creating an average performance for the all game can reduce the precision of the result. Strategies preference and risk aversion profile are also to be assessed for each gameplay. These three criteria need to analyze with attention the game system to define the gameplay pattern.

Efficiency and effectiveness assessment with precision requires clarifying what is the player's objective. In sand box games we need a layer of interpretation to define the objective.

Completion is the percentage of game completed within reachable content. The interactive objects and interaction potentials must be clearly defined for each game. It can be mapped to discover goals, weapon to use, a collection to complete, a quest to activate…The player must know it all.

Ambiguity needs to be assessed to measure what the player knows of a situation. The difficulty here is to understand how he perceives information. We can eventually trace all the critical elements displayed to him by sound, rumble and images. Even by doing so, we cannot be sure of his perception and understanding.

Transforming Motivation localization in measurable token in the game needs layers of interpretation and can create bias in the result. The simple fact you open a shop interface can mean the player perceives an expensive object that can drive his gameplay choices for a long period of time. Pay-off term preference is in the same case. For these two criteria the way we build the interpretation had to be exhaustively documented and communicated to the final user of the player profiling tools.

Some criteria can be measured with temporality influences as an historic or by mixing other parameters: performance progression through time, strategic preference of people at noon and during the evening.

4 Conclusion

The criteria table is the draft of an individual gameplay player profile grid. It can be adapted to several types of games. This plasticity is a progression from previous behavioral approach. We do not impose types; records of a population may reveal specific behaviors for a specific game.

Our next step will be to embody the grid in a game and perform real time recordings.

References

1. Bartle, R.A.: Hearts, Clubs, Diamonds, Spades: Players who suit MUDs. Journal of MUD research 1 (1996); reprint in Salen, K., Zimmerman, E. (ed.) The Game Design Reader. MIT Press, Cambridge (2006)
2. Juul, J.: Half-real. MIT Press, Cambridge (2005)
3. Henriot, J.: Sous couleur de jouer. Jose Corti Publisher, Paris (1989)
4. Huizinga, J.: Homo ludens, Essai sur la fonction sociale du jeu. Gallimard, Paris (1951)
5. Salen, K., Zimmerman, E.: Rule of Play: Game Design Fundamentals. MIT Press, London (2004)
6. Koster, R.: Theory of fun for game design. Paraglyph Press, Scottsdale (2005)

7. Guardiola, E.: Ecrire pour le jeu. Dixit, Paris (2000)
8. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper Perennial, New York (1991)
9. Bethesda: The Elder Scrolls IV Oblivion. 2K Games Publisher (2006)
10. Microsoft game studios: Forza Motorsport. Microsoft Game Studios Publisher (2005)
11. Herbrich, R., Graepel, T.: Playing machines: machine learning application in video games, `http://research.microsoft.com/mlp/Forza/FMD_manual.htm?0sr=a`, Applied game group, Microsoft Research Cambridge, `http://research.microsoft.com/en-us/projects/drivatar/forza.aspx`
12. Namco: Pacman (1979)
13. Taito: Space Invader (1978)
14. Tversky, A., Kahneman, D.: Advances in Prospect Theory: Cumulative Representation of Uncertainty. Journal of Risk and Uncertainty 5, 297–323 (1992)
15. Guardiola, E., Natkin, S.: Game Theory and video game, a new approach of game theory to analyze and conceive game systems. In: CGAMES 2005, Angouleme (2005)
16. Denis, G.: Jeux vidéo éducatifs et motivation: application à l'enseignement du jazz, Thèse à l'Ecole des mines de Paris (2006)
17. Deci, E.L., Ryan, R.M.: The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. Psychological Inquiry 11, 227–268 (2000)
18. Blizzard: World of Warcraft. Vivendi Universal Publisher (2004)
19. Mimesis Republic: Mamba Nation, in development. Mimesis Republic Publisher (2010)

A Laban-Based Approach to Emotional Motion Rendering for Human-Robot Interaction

Megumi Masuda, Shohei Kato, and Hidenori Itoh

Dept. of Computer Science and Engineering, Graduate School of Engineering,
Nagoya Institute of Technology,
Gokiso-cho Showa-ku, Nagoya 466-8555, Japan
{masuda,shohey}@juno.ics.nitech.ac.jp

Abstract. A motion-rendering system that adds target emotion to basic movements of human form robot (HFR) by modifying the movements was created. Pleasure, anger, sadness or relaxation is considered as target emotion. This method not only keeps the user interested, but it also makes the user perceive the robot's emotions and form an attachment to the robot more easily. An experiment was conducted using a real HFR to test how well our system adds target emotion to basic movements. The average of the success rates for adding the target emotion to basic motions were over 60%. This suggests that our method succeeded in adding the target emotions to arbitrary movements.

1 Introduction

We believe that communication robots will take an active part in our daily lives in the near future. There are many studies about communication robots (e.g., [1], [2]). Moreover, some communication robots are already in use.

Expression of emotions by robots is essential for human-robot communication (HRC). There are two methods for a robot to express an emotional state; one uses verbal information (e.g., [3], [4]), and the other uses nonverbal information (e.g., [5], [6]). In this paper, whole-body expression as nonverbal information is considered as an expression of emotion (e.g., [7], [8]).

Our aim is to create a motion-rendering system that adds target emotion to basic movements by modifying the movements. Canned whole-body movements get old fast because those movements have little variation. The system not only keeps the user interested with much variation, but it also makes the user perceive the robot's emotions and form an attachment to the robot more easily.

First, we introduce Laban movement analysis and Laban's feature value set, which we defined. Next, we describe the method for adding target emotion to arbitrary basic movements, which is based on the result of pilot experiment [9]. In this paper, we consider pleasure, anger, sadness or relaxation as a target emotion. Finally, an experiment is conducted to confirm the usefulness of our method.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 372–380, 2010.

Table 1. Laban's features

		Large	Small
Space	the movement directions are	one-sided	different
Time	whole-body movement is	quick	slow
Weight	whole-body movement is	strong	weak
Inclination	the whole-body is biased	forward	backward
Height	the whole-body is biased	upward	downward
Area	the range of whole-body is	large	confined

2 Correlation between Laban's Feature and Emotions Expressed by Movements

Previously, we proposed a set of motion feature values, called the Laban's feature value set, on the basis of Laban movement analysis (LMA) [9]. In this section, we discuss the features of LMA, and the correlations between robot's motion features and its emotions, estimated by an observer.

2.1 Laban Movement Analysis

LMA [10] is a well known theory which developed by Rudolf von Laban, who is widely regarded as a pioneer of European modern dance and a theorist of movement education. It has succeeded Darwin's movement theory [11], which focuses on the structure of an animal's bodily expression. Laban's theory is well suited for science and engineering, because it is mathematical and specific.

2.2 Laban's Features

The six main features of LMA are *Space*, *Time*, *Weight*, *Inclination*, *Height* and *Area*. We define Laban's features as follows and Table 1.

- ***Space*** represents the bias of whole-body movement.
- ***Time*** represents the quickness of whole-body movement.
- ***Weight*** represents the powerfulness of whole-body movement.
- ***Inclination*** represents the bias for forward of posture.
- ***Height*** represents the straightness of posture.
- ***Area*** represents the range of whole-body body.

Please refer to [9] for more precise.

2.3 Correlation between Laban's Features and Expressed Emotions

We have conducted a pilot experiment to examine the correlation between robot's motion features and the emotions expressed by movements [9]. Laban's feature value set is used as the motion feature values. KHR-2HV (Degree-of-freedom = 17, height = 353 mm, shown in Fig. 1) is used as HFR.

Table 2. Correlations between Laban's Features and Expressed Emotions

	Space	*Time*	*Weight*	*Inclination*	*Height*	*Area*
Pleasure	-0.04	0.45	0.46	-0.27	0.33	0.36
Anger	-0.21	0.30	0.33	0.01	-0.02	0.20
Sadness	0.03	-0.38	-0.42	0.47	-0.51	-0.39
Relaxation	0.16	-0.15	-0.12	-0.37	0.36	0.01

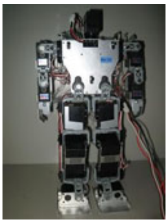


Fig. 1. KHR-2HV

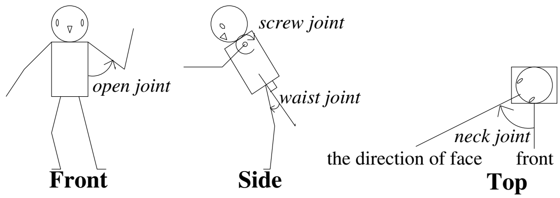


Fig. 2. Link Structure of KHR-2HV and Information

The results of the correlations are in Table 2. Light gray represents a positive correlation (significance level is over 1%), and dark gray represents a negative correlation (significance level is over 1%). This table suggests the following.

Pleasure correlates with quickness and powerfulness. It also correlates a bias for backwards posture, straightness of posture, and range of body.
Anger correlates with quickness and powerfulness. It also correlates with movement with moves in different directions, and range of body.
Sadness correlates with slowness and weakness. It also correlates with a bias for forward posture, low posture, and narrowness of body.
Relaxation correlates with slowness and weakness. It also correlates with movement in the same direction, a bias for backwards posture, and straightness of posture.

3 Method for Adding Emotion to an Arbitrary Basic Movement

In this section, we describe the method for adding target emotion to arbitrary basic movements. We created a motion rendering system with the consideration that movement can be emotive if it is processed on the basis of the correlation between a robot's motion features and its emotions (Table 2). In this paper, arbitrary basic movements are limited to movements that do not use the feet to reduce the risk of falling.

3.1 Method for Adding Emotion to Basic Movements

Our method, whose aim is to add target emotion to basic movements, processes basic movements to change the Laban's feature values on the basis of the correlation. In this paper, we adapt HFR KHR-2HV (shown in 1) as the agent. Fig. 2 is a link structure of the KHR-2HV and information about the method.

A basic movement is modified at every unit timepoint t as follows:

Space represents the bias of whole-body movement.
The bias of whole-body movement is related to the movement direction of the extremities and the direction of the face. The system modifies the direction of the face. If the direction of the face is near the average direction of the movement directions of extremities, we consider the directions of movement to be one-sided. The system modifies the direction of the face to add emotion as follows.

$$\theta_{head}(t) \leftarrow \theta_{head}(t) + (\theta_{max}(t) - \theta_{head}(t)) \times emo \tag{1}$$

θ_{head} is the angle of the face (*neck joint*) and θ_{max} is the average of all movement directions of extremities. The *emo* is the correlation coefficient between anger or relaxation and *Space*.

Time represents the quickness of whole-body movement.
The quickness is related to the angle velocities. The system modifies movement to be quicker. The system shortens the time necessary to rotate a certain angle to add emotion as follows.

$$time\_interval(t) \leftarrow time\_interval(t)(1-a) \tag{2}$$

The *time_interval* is the time necessary to rotate a certain angle. The a is the weight coefficient.

Weight represents the powerfulness of whole-body movement.
Powerfulness is related to angle accelerations. Powerfulness changes at linearly with *Time*,

$$\ddot{\theta}(t) \leftarrow \frac{\dot{\theta}(t+1)}{(1-a)} - \frac{\dot{\theta}(t)}{(1-a)} = \frac{(\dot{\theta}(t+1) - \dot{\theta}(t))}{(1-a)} = \frac{\ddot{\theta}(t)}{(1-a)} \tag{3}$$

Inclination represents the bias for forward posture.
The bias for forward posture is related to the center gravity of the body. The system makes the object bend forward by changing the angle of the waist The system modifies the angle of the waist to add emotion as follows.

$$\theta_{waist}(t) \leftarrow \theta_{waist}(t) + b \times emo \tag{4}$$

θ_{waist} is the angle of the waist (*waist joint*). The *emo* is the correlation coefficient between pleasure, sadness, or relaxation and *Inclination*. The b is the weight coefficient.

Height represents the straightness of posture.
The straightness of the posture is related to the center gravity of the body. The system raises both hands of the KHR-2HV. The system modifies the angle of the shoulder joint to add emotion as follows.

$$\theta_{shoulder}(t) \leftarrow \theta_{shoulder}(t) + c \times emo \tag{5}$$

$\theta_{shoulder}$ is the angle of the shoulder (*screw joint*). The *emo* is the correlation coefficient between pleasure, sadness, or relaxation and *Height*.

Area represents the range of the body.

The range of the body is related to the quadrilateral area, which is made by the four extremity points of the extremities, on the horizontal plane. The system makes both hands nearly horizontal. The system modifies the angle of the shoulder joint to add emotion as follows.

$$\theta_{shoulder2}(t) \leftarrow \theta_{shoulder2}(t) - (180 - \theta_{shoulder2}(t))emo \quad (when\ \theta_{shoulder2}(t)\ is\ over\ 90) \quad (6)$$
$$\theta_{shoulder2}(t) \leftarrow \theta_{shoulder2}(t) + \theta_{shoulder2}(t) \times emo \quad (when\ \theta_{shoulder2}(t)\ is\ under\ 90) \quad (7)$$

$\theta_{shoulder2}$ is the angle of the shoulder (*open joint*). $\theta_{shoulder2}$ is 90 when the arm turns to horizontal. The *emo* is the correlation coefficient between pleasure, anger, and sadness and *Area*.

4 Impression Assessment

We conduct an experiment to test the usefulness of our method. It is the experiment to see if people could identify the emotions we attempted to add to the robot's movements. There were twenty-one subjects between the ages of 20 and 40. They observed KHR-2HV's whole-body movements for about eight seconds and estimated its emotions. Three basic movements (shown in Fig. 3) were prepared. There were four processed movements for the each of the basic movements. We explained the experimental procedure to the subjects and conducted the experiment. The questionnaires used for the experiment are shown in Figs. 4 and 5. The procedure of the experiment is as follows.

1. Subjects watch the basic movement of KHR-2HV. The subjects estimate and mark how strongly they think each emotion is expressed through the basic movement. They marked the segment in the questionnaire in Fig. 4 to answer. The more strongly the subject perceives that the robot expresses the target emotion, the closer the mark should be to 1. The less emotion the subject perceives, the closer the mark should be to 0. Homogeneous transformation following the marking was done to quantify the estimation between 0 to 100. The averages of certain emotion's quantified estimations are called the "average of subject's estimations".

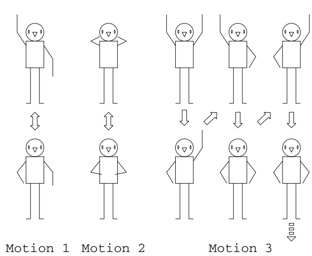


Fig. 3. Beckoning Motions

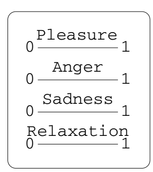


Fig. 4. Questionnaire for the basic movements

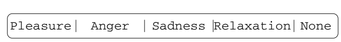


Fig. 5. Questionnaire for additional emotions

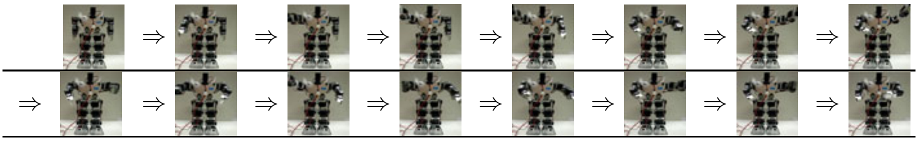


Fig. 6. An Example of Basic Movement (Motion 3)

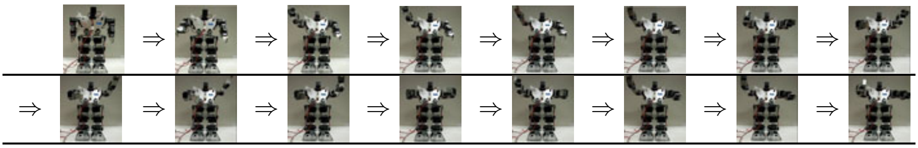


Fig. 7. An Example of Processed Movement (target emotion : Relaxation)

2. The subjects watch a processed movement. The subject evaluates what emotion is added. The evaluations are expressed by marking at the right point of Fig. 5.
3. The subjects watch the basic movement again.
4. Two and 3 are repeated until four processed movements run out.

We present any movement again if the subject requests it.

The above experiment was conducted three times with different basic movements. Because our aim is to add emotion to arbitrary movements, the three basic movements were made as movements giving different impressions. Fig. 3 shows these motions. For example, continuous snapshots of the processed movement of Motion 3 (shown in Fig. 6) is shown in Fig. 7.

5 The Rate of Success

The averages of subjects' estimations are shown in Table 3. The averages of subjects' estimations are between 0 and 100; bigger values mean the movement expresses the emotion more strongly. Table 3 suggests that the three basic movements give different impressions.

5.1 Rates of Success for Adding the Target Emotion

The rates of emotion felt by subjects, which is the rate of the subjects feeling certain emotion at certain processed movement, are shown in Table 4. This shows that intended emotions are best supported by the subjects.

The rates of success of adding emotion are shown in Table 5. These are the rate of the subjects identifying the targeted additional emotion.

First, we explain the rates of success of adding the target emotion for each emotion. All emotions had a high rate of success. The rate of success for adding sadness was especially high. Meanwhile the rate of success for adding sadness

Table 3. Averages of Subjects' Estimations of Basic Movements

	Motion1	Motion2	Motion3
Pleasure	39	16	68
Anger	30	13	30
Sadness	15	27	6
Relaxation	15	27	14

Table 4. Rates of emotion felt by subjects

target\felt	Pleasure	Anger	Sadness	Relaxation
Pleasure	58.7	29.4	1.6	4.8
Anger	34.9	49.2	1.6	2.4
Sadness	0.8	8.7	82.5	4.8
Relaxation	8.7	7.1	16.7	59.5

Table 5. Rates of Success of Adding Emotion

	Motion1	Motion2	Motion3	Average
Pleasure	64.3	54.8	57.1	58.7
Anger	73.8	31.0	42.9	49.2
Sadness	85.7	88.1	73.8	82.5
Relaxation	50.0	71.4	57.1	59.5
Average	68.5	61.3	57.7	62.5

to Motion 3 is comparatively low. We think the reason is that the strength of pleasure for Motion 3 is high (see Table 3). In this particular example, it is difficult to add sadness to Motion 3, because the basic movement expresses strong pleasure; the opposite of sadness in Russell's circumplex model [12]. The rate of success for adding anger was comparatively low. We think the reason is that there were few anger movements in pilot experiments. Therefore, it is probable that the motion features of anger were not distilled sufficiently.

Next, we explain the rates of success for adding the target emotion to each basic movement. All basic movements had high rates of success. The rate of success for adding the target emotion to Motion 3 is comparatively low. We think the reason is the lowness of the rate of success for adding sadness.

6 Related Works

There are several studies about distilling Laban's features. The robot used by Nakata et al. [13] had three joints and moves on wheels. In contrast, we proposed the method for distilling a Laban's feature value set on a robot that is an HFR with much more joints. An HFR is the type of robot better adapted to mental-like interaction. Maeda et al. [7] studied emotions detection from body movements. They used images of humans and robots. Their information retrieved from the images are comparatively less. In contrast, our Laban's feature value set is sensitive to whole-body movements. Moreover, we considered body movements by a real robot, KHR-2HV. It is natural from a lot of studies (e.g., [14], [15]) that a robot agent can create more positive impressions than a virtual agent.

There are some studies of selecting behavior suitable for situation (e.g., [16], [17]). Our system will make robot more natural by joining hands with these studies. Amaya et al. [18] introduced a model to generate emotional animation from neutral human motion. They deal with two emotions and two motion features In contrast, we deal with four emotions and six motion features based on Laban movement analysis.

7 Conclusion

We proposed the set of motion feature values, called the Laban's feature value set, on the basis of Laban movement analysis (LMA). We described a concrete method of modifying arbitrary basic movements to add emotions for real human form robot. We conducted an experiment to test if our motion rendering system could add target emotion to basic movements. The results suggest that our method added the target emotion to the movements.

There are some challenges. Our next aim is adding emotion at the intended strength. In this experiment, we used only three basic movements. In the future, we will dedicate to experiment with much more movements to confirm the usefulness of our method. Moreover, our final aim is to build a system that can create its own effectual movements automatically.

Acknowledgment

This work was supported in part by the Ministry of Education, Science, Sports and Culture, Grant–in–Aid for Scientific Research under grant #20700199.

References

1. Bennewitz, M., Faber, F., Joho, D., Behnke, S.: Fritz – A Humanoid Communication Robot. In: Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 1072–1077 (2007)
2. Mitsunaga, N., Miyashita, T., Ishiguro, H., Kogure, K., Hagita, N.: Robovie-IV: A Communication Robot Interacting with People Daily in an Office. In: IROS, pp. 5066–5072. IEEE, Los Alamitos (2006)
3. Breazeal, C.: Emotion and Sociable Humanoid Robots. International Journal of Human-Computer Studies 59, 119–155 (2002)
4. Hara, I., Asano, F., Asoh, H., Ogata, J., Ichimura, N., Kawai, Y., Kanehiro, F., Hirukawa, H., Yamamoto, K.: Robust Speech Interface Based on Audio and Video Information Fusion for Humanoid HRP-2. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004), pp. 2404–2410 (2004)
5. Brooks, A.G., Arkin, R.C.: Behavioral overlays for non-verbal communication expression on a humanoid robot. Autonomous Robots 22, 55–74 (2007)
6. Itoh, C., Kato, S., Itoh, H.: A Characterization of Sensitivity Communication Robots Based on Mood Transition. In: Ho, T.-B., Zhou, Z.-H. (eds.) PRICAI 2008. LNCS (LNAI), vol. 5351, pp. 959–964. Springer, Heidelberg (2008)
7. Maeda, Y., Tanabe, N.: Basic Study on Interactive Emotional Communication by Pet-type Robot. Transactions of the Society of Instrument and Control Engineers 42, 359–366 (2006) (in Japanese)

8. Hattori, M., Nakabo, Y., Tadokoro, S., Takamori, T., Yamada, K.: An analysis of the Bunraku puppet's motions based on the phase correspondence of the puppet's motions axis-for the generation of humanoid robots motions with fertile emotions. In: IEEE International Conference on Systems, Man, and Cybernetics, pp. 1041–1046 (1999)
9. Masuda, M., Kato, S., Itoh, H.: Emotion Detection from Body Motion of Human Form Robot Based on Laban Movement Analysis. In: Yang, J.-J., Yokoo, M., Ito, T., Jin, Z., Scerri, P. (eds.) PRIMA 2009. LNCS, vol. 5925, pp. 322–334. Springer, Heidelberg (2009)
10. Laban, R.V.: Mastery of Movement. Princeton Book Co. Pub. (1988)
11. Darwin, C.: On the Expression of the Emotions in Man and Animals. John Murray (1872)
12. Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology, 1161–1178 (1980)
13. Nakata, T., Mori, T., Sato, T.: Analysis of Impression of Robot Bodily Expression. Journal of Robotics and Mechatronics 14, 27–36 (2002)
14. Wainer, J., Feil-Seifer, D.J., Shell, D.A., Mataric, M.J.: Embodiment and Human-Robot Interaction. In: 16th IEEE International Conference on Robot & Human Interactive Communication, pp. 872–877 (2007)
15. Powers, A., Kiesler, S., Fussell, S., Torrey, C.: Comparing a computer agent with a humanoid robot. In: 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI 2007), pp. 145–152 (2007)
16. Sawada, T., Takagi, T., Fujita, M.: Behavior selection and motion modulation in emotionally grounded architecture for QRIO SDR-4X II. In: Proceedings of the IEEE International Conference on Intelligent Robots and Systems, vol. 3, pp. 2514–2519 (2004)
17. Kim, Y., Kim, Y., Kim, J.: Behavior selection and learning for synthetic character. In: Proc. of the IEEE Congress on Evolutionary Computation, pp. 898–903 (2004)
18. Amaya, K., Bruderlin, A., Calvert, T.: Emotion from motion. In: Proceedings of the conference on Graphics interface 1996, pp. 222–229. Canadian Information Processing Society (1996)

A Biofeedback Game with Physical Actions

Nagisa Munekata[1], Teruhisa Nakamura[2], Rei Tanaka[2], Yusuke Domon[2], Fumihiko Nakamura[1], and Hitoshi Matsubara[2]

[1] Hokkaido University, Sapporo, Japan
{munekata@complex.eng,fumihiko-n@ec}.hokudai.ac.jp
http://www.hokudai.ac.jp
[2] Future University-Hakodate, Hakodate, Japan
{b1007161,b1007216,b1007198,matsubar}@fun.ac.jp
http://www.fun.ac.jp

Abstract. We developed a biofeedback game in which players can take other physical actions besides simply “relaxing”. We used the skin conductance response for sensing a player’s surge of excitement and penalized players when they did not attack enemies in situations because they were not calm enough to meet the biofeedback threshold. We conducted a subjective experiment to to see whether people found the game enjoyable. Most participants felt the game was enjoyable.

Keywords: biofeedback, phisical actions, skin conductance response.

1 Introduction

Various forms of entertainment with biofeedback (BF) are being developed. This new trend is exemplified by Nintendo Ltd’s announcement that it was releasing a “Wii vitality sensor” that would measure players’ heart beat and other biological signals, as a peripheral device for its Wii game platform [1]. A game with BF has an unusual property whereby the player’s mental concentration is the key to victory or defeat [2,3]. The most popular video game series with BF is “The Journey to Wild Divine” [4]. The blurb on this game reads “Daily mind and body practice for healthy living,” and the game uses BF to make the user learn proper breathing and meditation techniques. In such games, the player completes game stages when he or she strives to relax through mental concentration. Most theorists endorse the view that emotions comprise three components: subjective experience (e.g., feeling joyous), expressive behavior (e.g., smiling), and a physiological component (e.g., sympathetic arousal); others add motivational state or action tendency and/or cognitive processing to that list [5]. However, most BF games have been regarded as merely instruments for the user’s health. That is, they are unsuitable as entertainment because if the user falls into a relaxed state while playing a video game they will become bored with it and eventually quit the game. Thus, BF has been viewed as a technology by which the user’s mind can perceives the body’s state in order to control it. BF is known for its sedating effect; it reduces stress or excitement and promotes relaxation.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 381–388, 2010.

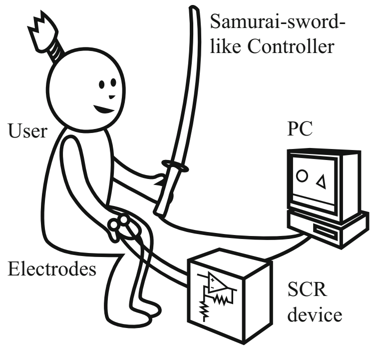


Fig. 1. System overview

Moreover, a common characteristic of BF games is that the player succeeds by keeping relaxed. Such a goal is not especially good for entertainment content, and thus far, BF's application to games has been limited. In contrast to the above passive uses, we have developed a number of BF games in users can be excited and agitated [6,7,8]. These games have the purpose of inducing "out-of-control" emotions such as the feeling of being on a tightrope. Our previous studies were dedicated to exploring the entertainment possibilities of BF. Many games, including those developed by us, however, require a player to relax in a sitting posture and maintain a resting state. The inputs to these games consist of only the emotional expressions mediated by BF as well as pushing of controller buttons and arrow keys. In this study, therefore, we developed a BF in which the user not only has to relax but also has to execute physical motions. Specifically, the user has to shake the hand controller to have an effect on the game. we thought the introduction of physical motions as a second form of input would excite user and enhance the quality of entertainment despite that user should stay relaxed. We evaluated this game in a subjective experiment that assessed the users' feelings of excitement.

2 System Configuration

The electrodes are placed on the user's hand to measures change in his or her skin conductance response (SCR) (see Fig. 1). The changes sensed by the SCR values reflect the user's excitement regarding the gameplay. We made the game harder for the player to win if these changes are large. Moreover, we made a samurai-sword-like controller as the way of inputting physical actions. The user inputs an action by shaking the controller strongly. The idea is for the user to hit "enemies" in the game by using this controller. The elements of the game system are detailed below.

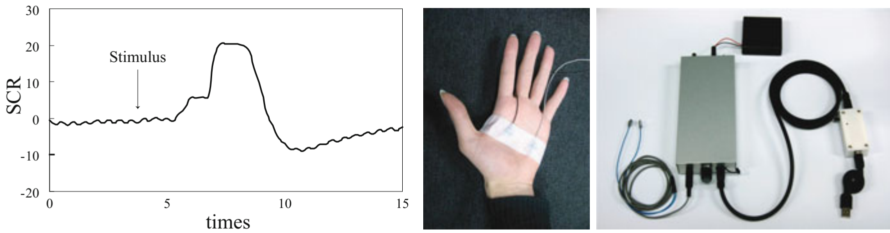


Fig. 2. Graph of SCR and device

2.1 Skin Conductance Response

Electrical signals from a living body have been used for medical diagnosis and treatment and in lie detectors used in police questioning [9]. The biological signal used in the lie detector is the skin conductance response (SCR); the conductance on the skin surface changes as a result of sweating induced by mental agitation, surprise, and excitement [10]. We have little awareness of the physiological functions of our own body because most physiological functions are involuntary, and therefore uncontrollable. The SCR is a typical example in this regard. No one is aware of the minute amounts of sweat that well up during mild mental agitation. Figure 2 shows an SCR graph, the measuring device and a user's hand with electrodes attached to it.

2.2 Samurai-Sword-Like Controller

The samurai-sword-like controller is a sword shaped toy with a built-in tri-accelerator. The system converts analog data from the tri-accelerator (AE-KXP84) into digital data by using a microprocessor (PIC12F675), and it is connected to a PC through a USB conversion module (AE-UM232R) and USB mini cable. We made a sword hilt with a built-in base (2×6 cm) that was soldered to the tri-accelerator, microprocessor, and USB conversion module. The USB mini cable can be unspooled from a small hole in the hilt and connected to the PC. When the user shakes this controller vigorously (acceleration: more than 0.20 m/s^2), he or she can hit enemies in the game. We set the standard of "shaking vigorously" to be the minimum acceleration measured during a test in which ten persons shook the controller vigorously.

2.3 Story of the Game

We chose a "fly swatter" game in which the player hits and kills flies. In this game, the user must restrain his or her excitement and attack with physical movements. Needless to say, fly swatting is not easy in real life. If you swat too much, the flies will notice the danger and fly away. As in real life, the player in the game must relax in order to succeed and he or she must attack quickly when the chance comes. There is a strong resemblance between the intended attitude

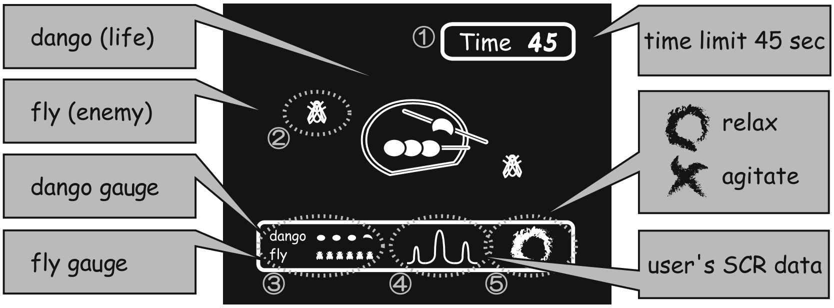


Fig. 3. Display of the game

and the purpose of this game, and that is why we thought the game would be a good BF application.

The flow of this game is as follows. First, the user watches a movie introducing the story behind the game. In this movie, a samurai is resting in a teahouse and eating dangos. (A dango is a Japanese dumpling made from rice flour.) A samurai finds flies on his dangos. This makes him angry and he tries to swat them with his sword. After the user finishes watching this movie, the game begins and the user attempts to hit the flies with his or her samurai-sword-like controller.

2.4 Game Rules

Figure 3 shows ① the time limit, ② flies, ③ dango gauge and fly gauge, ④ SCR graph, and ⑤ state signs of attack (agitating:✕, not agitating:○). ⑤ depends on "user's agitation" level as determined by the SCR value.

A user tries to hit flies on the screen until the time limit (45 seconds) expires. When the SCR values exceed the threshold value (70), the user is deemed to be agitated, and the "✕" graphic is displayed at position "⑤" on the screen. On the other hand, when the SCR does not exceed the threshold, the user is deemed to be relaxed and "○" graphic appears. The user can hit flies whenever he or she wants. However, if the user tries to hit them when the "✕" mark is displayed, the attack fails and the fly steals a dango. Thus, the game progresses poorly if the user attacks too often without paying attention to the BF and relaxing. In addition because of the time limit, a player cannot succeed without attacking quickly. The threshold value was taken from the results of a previous study[6,7,8]. Three flies on average appear during five seconds, and their appearance within is uniformly random. After the game is over, the user's samurai level is displayed. The score is calculated from the evaluation value of each parameter in Fig. 4. The purpose of showing the "level of samurai" is to enhance the quality of entertainment because it is easy to compare to one's own level with others. The parameters are described below.

– **Remaining flies:** the remaining number of flies in game (fly gauge): default value - twenty
– **Dangos remaining:** the remaining number of dangos in the game (dango gauge): default value - four

	User's state and judgement	Point
escapeFly	A fly is escaped from the display	-1
overThreshold	User's SCR is over the threshold	-5
Remaining dangos	User attack on a fly when user's SCR is over the threshold	-10
Display smached	Despite there are no flies, user attack on them three times	-10
brokenTeahouse	"crackDisplay" is occurred four times	-30
gameClear	All flies are exterminated	+30

Fig. 4. Detected user behaviors and corresponding points added or subtracted

We determine the "level of samurai" from the user's actions (see Fig. 4). Points are deducted and awarded as follows. **flyEscaped** means that the user attacks fly when the '✕' mark is displayed, the attack fails and the fly steals a dango. **overThreshold** means the user was excited and the SCR exceeded the threshold. **stolenDango** means that user attacked despite being **overThreshold**. At the same time, **Remaining dangos** is decremented. The game is over when **Remaining dangos** reaches zero. **Display smashed** means the user attacked even though flies were not on the screen. Such an act is considered very inappropriate of a samurai, and it thus entails a penalty of -30 points. **brokenTeahouse** means **Display smashed** occurred four times. If the user breaks the teahouse, the game is over and the user did not successfully complete the game. **gameClear** means that **Remaining flies** is zero and all of the flies were swatted. It entails a 30 point bonus.

We made 100 points the highest score and 0 the lowest. Moreover, this game is completed successfully when **Remaining flies**(fly gauge) reaches zero. The game finishes unsuccessfully when the **Time limit** is reached, **Remaining dangos** is zero (all of dangos were stolen by flies) or the teahouse becomes broken **brokenTeahouse**.

3 Experiment

The purpose of this experiment was to determine whether users would enjoy playing this BF game. So we investigated how BF affected the user in this game by comparing the SCR values with and without BF.

3.1 Experimental Method

Nine university students (eight men and one woman: 20-23 years old) participated in the experiment. As the dummy indicator, the average of randomly chosen SCR values of all participants was displayed as the SCR graph (Fig.3: ④), and this average was used to determine the state signals of attack (✕ or ○) (Fig.3 : ⑤). Before starting the experiment, participants watched a movie (about three minutes long) on how to play. Participants played the game with BF (phase 1) and without BF (phase 2) for three turns in the same order. They

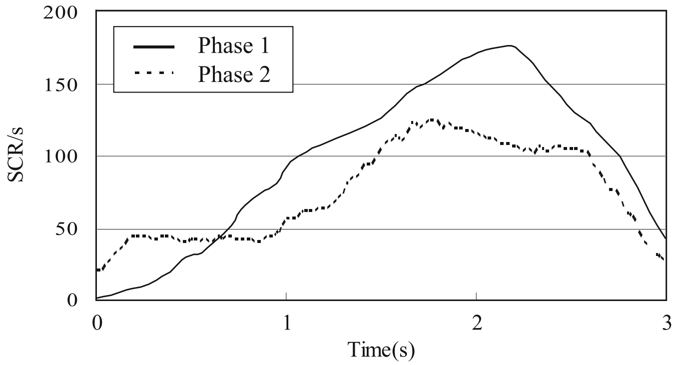


Fig. 5. Experimental results

were not informed of the distinction between these games. At the end of the experiment, the players were interviewed about whether they found the game enjoyable. Participants were randomly assigned to the following two groups:

– **Group A:** Participants undertook three trials for phase 1 and three trials for phase 2, and
– **Group B:** Participants undertook trial reversed phase order, i.e., phase 2 then phase 1.

3.2 Results

The log data of phase 1 and phase 2 were analyzed (Fig.5). In Fig. 5, SCR/s was defined as the SCR values per unit time and the extracted range was between the moment that the participant attempted to hit the fly and three seconds later. The results of a dispersion analysis showed significant differences ($F(1, 202) = 26.20, p < 0.01$) between group A and group B. This indicated the game with BF (phase 1) could make the user significantly excited more than the game without BF (phase 2.) Additionally, none of the participants notice that different games were played or that phase 1 was different from phase 2.

3.3 Discussion

The participants' excitement during phase 2 was significantly lower than in phase 1. This indicated that not only the SCR graph (Fig.3 : ④) and the state signals of attack (✕ or ○) (Fig.3 : ⑤) on the game screen increased excitement and agitation but also that the user's SCR value was the most important element contributing to excitement. Although the tests were conducted only six times, the results seem to indicate that had they played this game more, it would become easier for them to distinguish the phases 1. Participants affected their SCR values naturally without feigning even though they did not notice these distinctions. Therefore, we can conclude that the system exploited the characteristics of BF effectively to raise the users' level of excitement.

Let us consider the meaning of the SCR graph and the state signals of attack (Fig.3). People usually go about their daily business with few chances to assess their behavior on the basis of physiological signals. However, they may become partially aware of certain states of the body and mind without using equipment. For example, a rapid heartbeat may signal one of many states and user might try to calm his or herself by breathing deeply. Other people might observe that sweating from palms indicates a state of anxious suspense. These signals are natural responses, and there are cases in which the reason for them can be revealed by tracing memories even when the reason for physical changes cannot be quickly recognized. People always try to take the optimum actions to fit the situation by observing and altering their own states. Participants who played the game in this study also tried to make their state ideal for the game, for instance, by deep breathing, humming or easing the tension in their shoulders. However, these actions tended to increase the participants' stress or agitation level.

In fact, it is difficult to recreate the critical conditions under which the user would feel themselves in danger in a game. However, as evidenced by the significant autonomic nervous system activity of palms sweating and the participants' responses, we believe that a game involving awareness of physical signals in the form of biofeedback would have good entertainment value.

4 Conclusion

We created a new BF game that stimulates the player's sense of agitation. This game requires the player not only to relax but also to perform physical actions. Previous BF games have had elements whereby the user's mental concentration makes the situation advantageous and his or her agitation makes the situation disadvantageous. We added a new element: i.e., an attacking action that involves the player shaking a controller with one hand. We experimented with this BF game to examine whether the players would feel more agitation when the BF was used and in turn feel more entertained.

The experimental result showed that the BF game was entertaining and that the individual variation in the SCR values had very little influence on the game's outcome. In addition, the experimental results showed that the sources of the user's agitation came from not only making attacks but also from a situation in which he or she had to inhibit their level of agitation. Moreover, we found that feeding back the user's own SCR values in real-time helped to increase their level of agitation. Additionally, the participants in the experiment intuitively understood that the fly could not be hit when they themselves were agitated. The clear influence of the SCR values on the game reinforced the idea that to make the attack succeed, calm judgment is needed.

In past BF games, the influence of individual variations was a problem that needed to be solved. In this game, the problem was solved by involving BF in the user's attack behavior. Users showing a wide range of individual variations in SCR found the BF game to be entertaining. We added a second input method (physical action) that past BF games did not have. This input method agitated the user and improved the entertainment value, without deviating from the concept of a BF game. Until now, an entertainment application of BF had been assumed to be problematic. Our experimental results, in contrast, indicate that BF may be used for entertainment purposes in the future.

We will further the possibilities of BF based on this game and the experimental results obtained in this study. We wish to clarify the prospects of using BF in various fields, by developing an affective system that generates appropriate events for the user while monitoring his or her physiological state.

References

1. Nintendo Co., Ltd., http://www.nintendo.co.jp/index.html
2. Breathing Space, Media Lab Europe (2003)
3. Hand-held Doctor. MIT Media Laboratory (1997)
4. The Wild Divine Project, http://www.wilddivine.com
5. Ravaja, N., Saari, T., Laarni, J., Kallinen, K., Salminen, M., Holopainen, J., Jarvinen, A.: The Psychophysiology of Video Gaming: Phasic Emotional Responses to Game Events. In: Proceedings of DiGRA 2005 Conference: Changing Views-Worlds in Play, 16–20, 2005, pp. 1–13 (2005)
6. Sakurazawa, S., Yoshida, N., Munekata, N., Omi, A., Takeshima, H., Koto, H., Gentsu, K., Kimura, K., Kawamura, K., Miyamoto, M., Arima, R., Mori, T., Sekiya, T., Furukawa, T., Hashimoto, Y., Numata, H., Akita, J., Tsukahara, Y.: A Computer Game Using Galvanic Skin Response. In: 2nd IFIP International Conference on Entertainment Computing, pp. 31–35. Carnegie Mellon University, Pittsburgh (2003)
7. Sakurazawa, S., Yoshida, N., Munekata, N., Tsukahara, Y., Matsubara, H.: Entertainment Feature of the Computer Game Using a Skin Conductance Response. In: ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, pp. 181–186. ACM Press, New York (2004)
8. Munekata, N., Yoshida, N., Sakurazawa, S., Tsukahara, Y., Matsubara, H.: Design of positive biofeedback using a robot's behaviors as motion media. In: Harper, R., Rauterberg, M., Combetto, M. (eds.) ICEC 2006. LNCS, vol. 4161, pp. 340–349. Springer, Heidelberg (2006)
9. Geddes, L.A.: History of the Polygraph, an Instrument for the Detection of Deception. Bio-med. Eng. 8, 154–156 (1973)
10. Gross, J.J., Levenson, R.W.: Hiding feeling, the acute effects of inhibiting negative and positive emotion. Journal of Abnormal Psychology 106, 95–103 (1997)

Dial-Based Game Interface with Multi-modal Feedback

Wanjoo Park, Laehyun Kim, Hyunchul Cho, and Sehyung Park

Intelligence & Interaction Research Center,
Korea Institute of Science and Technology, Seoul, Korea
{wanjoo,laehyunk,hccho,sehyung}@kist.re.kr

Abstract. This paper introduces a dial-based haptic interface for a brickout game. Conventionally brickout games are played through a mouse or a keyboard. However, these input devices cannot provide sufficient intuitive interface to move the game paddle or provide multi-modal feedback for the user. We developed a haptic game device that gives the user haptic feedback during the game as well as visual and sound feedback. The user can move the position of the paddle by spinning the dial knob and feels various multi-modal effects according to the game context. Basic haptic effects include detent, vibration, friction and barrier. We can generate any combinations of these effects by adjusting the amount, frequency, and direction of torque along the rotational path. The result of a user-study showed that the proposed haptic dial interface made a simple brickout game more fun and more interesting. Additionally, the users were able to focus on the game more easily than when they played using a mouse.

Keywords: haptic interface, brickout game, dial knob, tactile feedback.

1 Introduction

Recently haptic interfaces have been developed that allow users to interact with digital information via the sense of touch. Haptic senses include force/tactile feedback, pressure, vibration, heat, and even pain. Haptic technology has various applications such as medical simulations [1], rehabilitation [2], mobile phones, automobiles, and games. Especially, the success of the Wii [3] and the the haptic phone [4] in providing tactile feedback has led to the integration of a haptic interface into many game devices. Most game devices such as Wii, PSP, and Nintendo DS give simple vibration effects to the user. PC-based video games are played by a mouse, a keyboard, or a joystick without tactile feedback.

Over the past several years, there have been a number of studies of haptic feedback controllers. Karon E. MacLean et al. introduced a variety of haptic devices and design parameters [5-7]. They considered characteristics of the touch sense and designed various tactile signals on their experimental devices. Several studies have been conducted of the 1 DOF (Degree of Freedom) dial knob. Scott S. Snibbe et al. suggested several haptic techniques for manipulating digital media based on intuitive physical metaphors [8]. Mircea Badescu et al. developed a single force feedback knob to imitate the senses of various conventional control knobs in motor vehicles [9].

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 389–396, 2010.

Laehyun Kim et al. used a haptic dial system for multimodal prototyping in the early phase of product development [10]. They applied the haptic technology to prototype the dial module of a washing machine.

There are several works to apply the haptic interface to games. Yuichiro Sekiguchi applied the haptic effect to a game, proposing a device that gives a user the illusion of a virtual object inside the device when shaking it using accelerators and actuators [11]. Jukka Linjama et al. developed a bouncing ball game that uses accelerator sensors and tactile feedback actuators [12]. Wanjoo Park et al. proposed a brickout game using a haptic dial interface and some haptic effects for the game [13].

2 Brickout Game and Hardware Configuration

The main objective of the brickout game is to remove all the bricks from the wall by hitting them with the ball (the right side of Fig. 1). In our brickout game, conventional bricks are replaced by banana bricks, a game ball is replaced by a monkey, and a paddle is replaced by people holding up boards. In addition, a new item is added, a cloth wrapper.

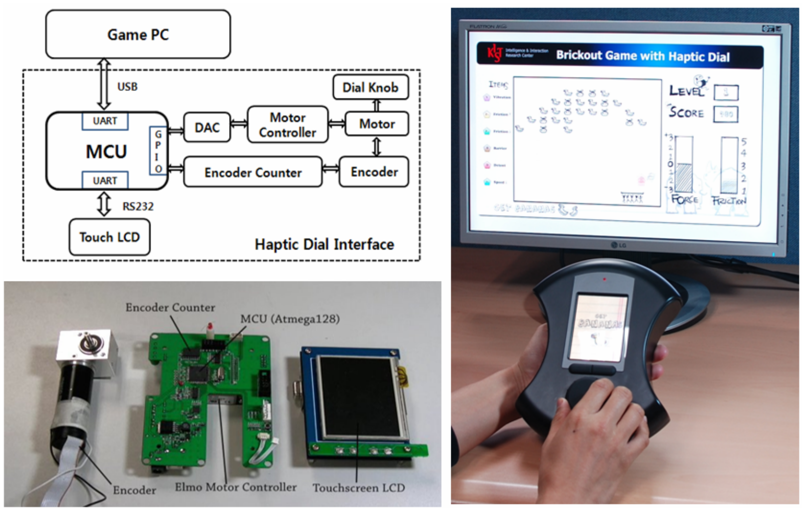


Fig. 1. Block diagram of the dial-based haptic system (*left upper side*), Modules of motor, control unit PCB and touch screen (*left down side*), Implementation of a brickout game (*right side*)

When the ball hits a cloth wrapper, the haptic item falls down. If the user catches the item by moving the paddle, a predefined tactile effect is felt that lasts a few seconds. The game's visual information helps the user understand the current game context.

We will explain in detail the hardware configuration of the haptic dial knob. A block diagram of the system is shown in the left upper side of Fig. 1. The main processor is an AVR ATMEGA-128 processor and the DAC (Digital to Analog Convertor) is a DAC0800 we obtained from National Semiconductor. A DC motor is used to generate various haptic patterns. We use an RE25 with a gear box of 5:1 ratio obtained from Maxon motor. The motor works at 24V/0.6A and 28.8mNm torque. The Dial knob is installed on the motor gear box, allowing the user to rotate the knob and to feel various haptic effects which are programmed along the angular position. The Encoder to measure the angular position is an MR 128 with 1000 pulse per turn from Maxon Motor. The encoder counter is an LS7166 manufactured by LSI Computer Systems. It has a 24-bits quadrature counter and a DC set to 25MHz count frequency. A LCD Module (an EZ-TFT350T manufactured by Alls Technology) is used to display visual information. It displays 3.5 inch QVGA 320 x 240 pixels with 24 bits true color.

When the user rotates the dial, the encoder generates the pulses. Then the encoder counter IC counts the pulses and the AVR processor reads the encoder value. The processor calculates the angle and drives the motor to generate haptic dial effects described in section 3. The user feels the programmed haptic effects according to the game context.

3 Haptic Dial Effects

In this section, we describe how to implement various haptic effects for a brickout game. Basic haptic dial effects include detent, vibration, friction, and barrier.

3.1 Detent Effect

The detent effect simulates small notches or a rugged surface along the angular position using sine functions. We call it the angle-based effect since the motor's torque function is implemented based on the dial's rotation angle. Eq. (1) shows the detent effect which is defined by a sine function.

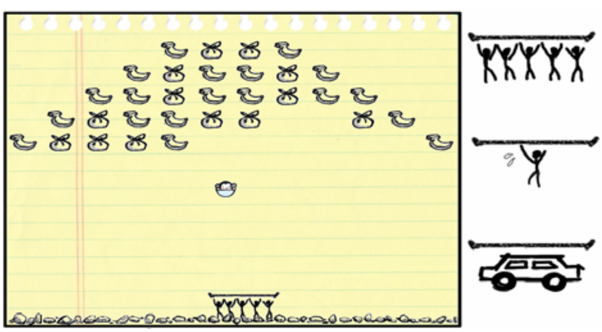


Fig. 2. Detent haptic effect (*left side*), these are paddle images for the different friction types *right side)*: Normal condition (*right upper side*), high friction (*right middle side*), low friction (*right down side*).

$$f_d(\theta) = A_\theta \sin(b \cdot \theta) \tag{1}$$

where, A_θ is the amplitude, b is the number of notches per turn, and θ is the rotation angle of the dial knob.

The torque profile for the detent effect is shown in the left side of Fig. 2. When a game player gets a detent haptic item, a gravel image appears under the paddle and an audio clip of walking on a gravel road is heard. The user senses the rugged surface through the multimodal stimulation of visual, aural, and haptic feedback. This effect makes control more difficult but makes the game more interesting and allows the player to focus on the game with a realistic physical feeling.

3.2 Vibration Effect

The vibration effects generate vibrotactile feedback as a time-based effect. This is represented by Eq. (2). We designed paddle vibrations and item vibrations. The paddle vibration has a small amplitude (A_v) and high frequency (s). When the paddle hits the game ball, the user feels the impact between the paddle and the ball according to the paddle's vibration. During this effect, the dial knob shakes weakly and rapidly and an audio clip of bouncing strings is played. This gives the user the sense that the ball is actually bouncing physically.

On the other hand, the item vibration is implemented by a large amplitude and low frequency. This effect generates a larger torque and longer period than the paddle vibration effect. When the user takes the cloth wrapper containing the item vibration effect, he/she feels several big vibrations lasting several seconds. As a result, paddle control becomes more difficult. For audio feedback, an audio clip of a warning siren adds tension for the gamer.

$$f_v(t) = A_v \sin(2\pi \cdot s \cdot t) \tag{2}$$

where, A_t is the amplitude constant, s is the frequency scaling constant, and t is time

3.3 Friction Effect

The friction effects generate resistant torque opposite to the direction of movement as a movement-based effect. This is implemented based on the friction con model [14]. When the dial knob is rotated slowly, the rotational friction is low. On the other hand, when the user spins the dial fast, he/she feels more strong friction. There are 3 different types of friction mode. The first is the friction mode for normal conditions. In this mode, the friction level L_f in Eq. (3) is small and the player feels relatively weak friction force. The paddle is represented by five walking people holding up a board. For the audio effect, an audio clip of many noisy people is used. The right side images of Fig. 3 show the paddle images for the different friction types. The second friction mode is the high friction effect. The paddle's image is transferred to only one person for the visual effect. An audio clip of the sound of a single person walking is used for the audio effect. Rotational friction force of dial knob is very high due to high L_f in Eq. (3). Accordingly, the player feels strong friction force as he/she moves the paddle. The third friction mode is a low friction effect. The paddle is transferred to a vehicle image and an audio clip of moving vehicle is used. In addition, the rotation friction of dial knob is near zero because of L_f is near zero in Eq. (3). Then the player can control the paddle's movement very easily and quickly.

$$\begin{aligned}
&P_{curr_f}(n) = P_{prev}(n-1) + (P_{now}(n) - P_{pre}(n-1)) \cdot S_f \\
&P_{diff}(n) = (P_{now}(n) - P_{curr_f}(n)) \cdot L_f \\
&P_{prev}(n) = P_{curr_f}(n) \\
&\textit{if} \quad P_{diff}(n) > T_{f_\max} \quad \textit{then} \quad T_f(n) = T_{f_\max} \\
&\textit{else if} \quad P_{diff}(n) < T_{f_\min} \quad \textit{then} \quad T_f(n) = T_{f_\min} \\
&\textit{else} \quad T_f(n) = P_{diff}(n)
\end{aligned} \tag{3}$$

where P_{curr_f} is the current position, P_{now} is the angular position of the dial knob, P_{pre} is the previous position, S_f is the scaling factor, P_{diff} is the difference of position, L_f is the friction level, and T_f is the friction torque.

3.4 Barrier Effect

The barrier effects simulate the hard stop as location-based effects because the barrier effect restricts the movement of the paddle in a certain location. Eq. (4) is a second order equation to calculate the barrier torque function. There are 2 types of barriers. In the normal condition, the user feels the barriers around the left and right wall. The barrier effect for the normal condition is calculated by $b=180^o$, $a=1$ and $w=6$. When b=180° it means that barriers are located on -180 °and 180 °. Variables of a and w in Eq. (4) are used to determine an edge function. With this, the player can feel a paddle cling to the wall like iron clings to magnet on near edge zones. If the player rotates the dial knob over the b point, the barrier effect blocks this rotation. The item barrier effect restricts the movement of a paddle inside a certain range for a certain time. In this case, b in Eq. (4) is set to 90, half of the normal condition. This effect is the most challenging item. If the game ball falls into the limited zone, the user cannot move the paddle inside that zone.

$$\begin{aligned}
f_b(\theta) &= \frac{-(\theta + b - w)^2}{w^2} + a, \quad (\theta \le -b + 2w) \\
&= 0, \quad (-b + 2w < \theta < b - 2w) \\
&= \frac{(\theta - b + w)^2}{w^2} - a, \quad (b - 2w \le \theta)
\end{aligned} \tag{4}$$

where θ is the rotation angle of the dial knob, b is the start point of barrier, a is the amplitude of edge function and w is the range of edge function.

5 User-Study

We performed a user study to compare the dial-based haptic interface with a mouse interface when playing the brickout game and to evaluate the interface's usability by using a heuristic evaluation method.

At the beginning, 130 words related to game evaluation were collected and then representative words were chosen by 4 experts. The similarities between the words were analyzed by cluster analysis. The cluster analysis is a statistic method of a common technique for statistical data analysis used in many fields. It is the assignment of a set of observations into subsets-clusters so that observations in the same cluster are similar in some sense. Finally 4 representative words were selected such as interesting, immersed, easy to control and fatigued (Table 1).

A usability survey was conducted with 35 undergraduate students regarding the 4 classified items. The survey processed in 2 steps. First, a volunteer was asked to play the brickout game with a mouse for 5 minutes, and to fill out the questionnaire. Additionally, the volunteer was asked to play the game with the dial-based haptic device for another 5 minutes and to answer the other questionnaire. The questions on the questionnaire were 5-point Likert scales for 4 adjectives: 'Interesting', 'Immersed', 'Easy to control', and 'Fatigued' (Table 1) and the volunteers gave the 4 adjectives scores from 1 (low) to 5 (high). A Likert scale is a psychometric scale commonly used in questionnaires, and is the most widely used scale in survey research. When responding to a Likert questionnaire item, respondents specify their level of agreement to a statement.

Table 1. Result of cluster analysis

Interesting	Immersed	Easy to control	Fatigued
fascinating	concentrated	easy	tired
absorbing	focused	simple	exhausted
attractive	centralized	freedom	weary
exciting	absorbed	efficiency of use	strenuous
fun	passionate	flexibility of use	arduous
pleasant	engrossed	clarity of goals	stressor
interesting	engaged	easy to use	fatigued
curiosity		intuitive	obstacles
wonder		easy to control	tension
fooling		casual	

Table 2. Result of T-test

	F	p-value	T
Interesting	8.9711	0.0038*	-9.0440
Immersed	4.3130	0.0416*	-5.8874
Easy to control	0.2710	0.6044	5.0556
Fatigued	0.5828	0.4478	-4.6697

* $P < 0.05$

The questionnaire was analyzed by *t*-test and numerical statement. A *t*-test is any statistical hypothesis test in which the test statistic follows a Student's *t* distribution if the null hypothesis is true. It is most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known. When the scaling term is unknown and is replaced by an estimate based on the data, the test statistic follows a Student's *t* distribution.

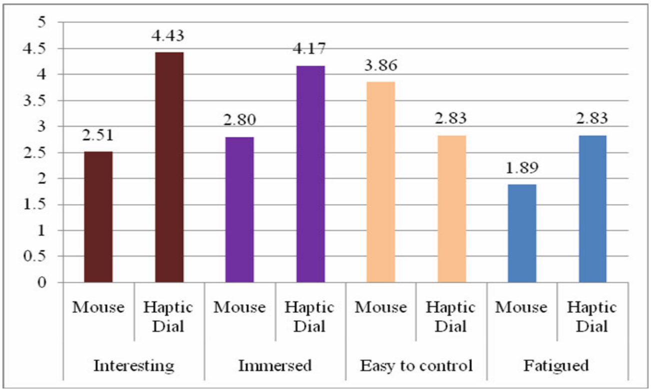


Fig. 3. Mean value of questionnaire

In Fig. 3, each bar indicates the average value of each item. The results show that the users felt more interest and more immersed in the game using the haptic interface than the mouse. However, the users could control the game paddle more easily with the mouse and felt somewhat more fatigued when playing with the haptic dial interface than with the mouse interface. According to a *t*-test (Table 2), though, it was concluded that the haptics dial control was more interesting and players felt more immersed in the game than when using the mouse control ($P<0.05$). It is less meaningful that the mouse interface is easier to control and less fatiguing because of the high p-value. Therefore, it is concluded that the game using the dial-based haptic interface is more interesting and immersed than with mouse interface. It is recommended that a dial knob be designed with an ergomechanic grip that is easier to control and less fatiguing.

6 Conclusion

This paper describes a dial-based haptic interface and various haptic effects for a brickout game. Conventional game devices such as Wii and PSP provide simple effects like vibration. However, the haptic effects we proposed give the user more active and various physical experiences. We designed several different haptic effects which have different amplitude, direction, and frequency of torque along the angular position or movement. These include detent, vibration, friction, barrier, slow motion, and combinations of these effects. The barrier effect can block the movement of the dial at specific points and the detent effect makes the user feel a rugged surface while moving the game paddle. These haptic effects, which can be used as penalties or rewards, can make the game more fun. In addition, the dial knob gives a very intuitive interface to change the position of the paddle in the brickout game compared with conventional input devices such as a keyboard, button, or mouse. We conducted a user-study to compare a conventional mouse interface with our newly developed dial-based haptic interface. According to the results of the survey, the haptic interface made the game more fun and helped the game player focus on the game.

In the future we will work on improving the haptic interface to reduce fatigue and facilitate greater control by designing an ergomechanic grip.

References

1. Arbabtafti, M., et al.: Haptic and Visual Rendering of Virtual Bone Surgery: A Physically Realistic Voxel-Based Approach. In: IEEE International Workshop on Haptic Audio Visual Environments and their Applications (HAVE 2008), pp. 30–35 (2008)
2. Alamri, A., et al.: Haptic Virtual Rehabilitation Exercises for Poststroke Diagnosis. IEEE Transactions on Instrumentation and Measurement 57(9), 1876–1884
3. Wii, N.: `http://www.nintendo.com/wii`
4. Behold Mobile Phone by Samsung Electronics, `http://www.samsungmobileusa.com/Behold.aspx`
5. MacLean, K.E.: Designing with Haptic Feedback. In: Symposium on Haptic Feedback in the Proc. of IEEE Robotics and Automation (ICRA 2000), pp. 22–28 (2000)
6. Hayward, V., Maclean, K.E.: Do It Yourself Haptics: Part 1. IEEE Robotics & Automation Magazine, 88–104 (December 2007)
7. MacLean, K.E.: Foundations of Transparency in Tactile Information Design. IEEE Transaction on Haptics 1(2), 84–95 (2008)
8. Snibbe, S.S., et al.: Haptic Techniques for Media Control. In: Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology (UIST 2001), pp. 199–208 (2001)
9. Badescu, M., Wampler, C., Mavroidis, C.: Rotary Haptic Knob for Vehicular Instrument Controls. In: Haptics 2002, Haptic Interfaces For Virtual Envir. & Teleoperator Systs., pp. 342–343 (2002)
10. Kim, L., Han, M., Shin, S.K., Park, S.H.: A Haptic Dial System for Multimodal Prototyping. In: 18th International Conference on Artificial Reality and Telexistence, ICAT 2008 (2008)
11. Sekiguchi, Y., Hirota, K., Hirose, M.: The Design and Implementation of Ubiquitous Haptic Device. In: First Joint Eurohaptics Conference, 2005 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, World Haptics 2005., pp. 527–528 (2005)
12. Linjama, J., Kaaresoja, T.: Novel Minimalist Haptic Gesture Interaction for Mobile Devices. In: NordiCHI 2004, ACM, pp. 457–458 (2004)
13. Park, W., Kim, L., Cho, H., Park, S.: Design of Haptic Interface for Brickout Game. In: IEEE International Workshop on Haptic Audio Visual Environments and Games (HAVE 2009), pp. 64–68 (2009)
14. Melder, N., Harwin, W.S.: Extending the Friction Cone Algorithm for Arbitrary Polygon Based Haptic Objects. In: Proceedings of the 12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS 2004), pp. 234–241 (2004)

Tangible Interactive Art Using Marker Tracking in Front Projection Environment: The Face Cube

Chan-young Bang[1], Jin Lee[2], Hye-won Jung[2], Ok-young Choi[2], and Jong-Il Park[1]

[1] Division of Electrical and Computer Engineering, Hanyang University,
17 Haengdang-dong, Seongdong-gu, Seoul, Korea
peap18@mr.hanyang.ac.kr, jipark@hanyang.ac.kr
[2] Division of Digital Media, Ewha Womans University,
11-1 Daehyun-dong, Seodaemun-gu, Seoul, Korea
mint__@naver.com, wint01@naver.com, okybom@hanmail.net

Abstract. The Face Cube is a work of interactive art which is targeted to children. To implement this art work, we use a camera-projector system. Instead of rear projection and edge detection method, we choose front projection approach and histogram-based detection method for interaction. This paper describes how to design the Face Cube and marker design for robust interaction, efficient way to remove projection lights from the front projection system for marker recognition, histogram-based marker detection, marker information management.

Keywords: the Face Cube, Interactive art, Marker tracking.

1 Introduction

The Face Cube was planned to exhibit at "Four Faces Exhibition", theme exhibition for children, by artist-technician collaboration. It was focused on intelligent exhibition space and interactive contents for children as a kind of interactive art.

Since 2005, the reacTable [1] has shown us new electronic musical instrument. In 2006, the Tablescape plus [2], Tokyo University, showed us interactive works with upstanding tiny displays on a tabletop display. Both famous interactive art-works use camera-projector system for interaction. They chose rear projection method and edge detection method for stable tracking. The rear projection method is useful in low ambient illumination settings. However, we didn't want gloomy space and hoped children would feel rays from projector delightfully like sunshine. Therefore, the Face Cube chose front projection method. Since we didn't want outlined markers, we chose histogram-based marker detection method for interaction.

This paper describes how to design the Face Cube and marker design for robust interaction, efficient way to remove projection lights from the front projection system for marker recognition, histogram-based marker detection, marker information management and artist-technician collaboration results.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 397–404, 2010.

2 Face Cube

The face cube was designed for "Four Faces Exhibition", as an effort for tight collaboration among engineers and artists. The Mixed Reality Laboratory in Hanyang University took the charge of technical parts and the Mobile and Tangible Media Laboratory in Ewha Womans University took the artistic parts.

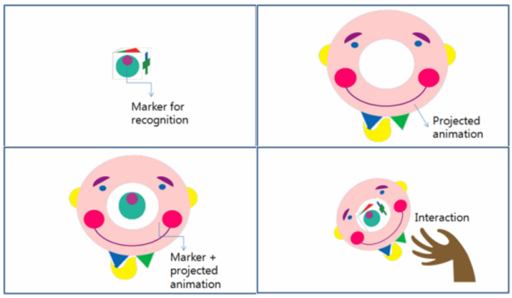


Fig. 1. Basic concept of the Face Cube

The Face Cube is a work of interactive art which is targeted for children. Children may roll a cube on the table. When a marker presented on each side of the cube is recognized by the camera, corresponding animation will appear around the cube. The marker is used not only to measure position or rotation of marker but also as a part of the animation output. Therefore, unlike using a square marker, we can reduce the visual unfamiliarity. Children can interact with and enjoy the artwork by encountering various unexpected characters.

3 Method

3.1 Artists and Technicians' Collaboration

The process of producing most works involves collaboration between technicians and artists beforehand. Some ideas from the artists may be modified due to the limits of techniques while some technical challenges could be restricted by the artists' request.

3.2 Marker Design and Detection Method

At first, technicians suggested marker detection method by drawing square-outlined markers around the cube. However, artists didn't easily accept the idea of square-outlined markers and they suggested symmetric markers but those were not suitable for finding direction. Through some iterative procedure between art and technology, we redesigned the markers as shown in Fig. 2.

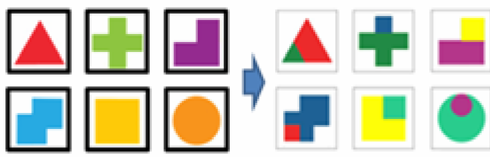


Fig. 2. First designed markers (left) and final markers (right)

The template matching method is usually used as a simple method to track objects or patterns. In our case, template matching is not suitable for detecting direction of markers, since we use symmetric markers of six different shapes. Markers are too simple to have feature points and thus feature point-based tracking methods, such as openSURF, are not adequate for this kind of tracking. Our choice was to use color histogram-based searching method which uses two colors to represent direction.

3.3 Dealing with Projected Images

The underlying equations of projected images can be represented as follows:

$$I_L = I_C + I_E + I_P(r, g, b). \qquad (1)$$

$$I_C + I_E < I_C + I_E + I_P(0, 0, 0) \leq I_C + I_E + I_P(r, g, b). \qquad (2)$$

where I_L is the camera image, I_C is the color of an equivalent flat frontal surface viewed under the same light, I_E is the influences of the environment lighting, and I_P is the influences of projector light (r, g, b means each values in RGB color space).

We need I_C to recognize the color of the cube, but it is difficult to fix one because of the environment lighting and projector light. Furthermore, the projector always displays black color, we call $I_P(0, 0, 0)$, on the cube. So the color on the cube during the detection includes $I_P(0, 0, 0)$. But images from the camera include $I_P(r, g, b)$. If the color of the cube, $I_C + I_E + I_P(0, 0, 0)$, has similar information to the color of the projected image, $I_P(r, g, b)$, it is difficult to see the difference between them. Thus if we know the color of the white table including $I_P(0, 0, 0)$, we may remove the color influences from the projector light by thresholding.

$$I_P'(r, g, b) = I_P(r, g, b) - I_P(0, 0, 0),$$

$$I_L = I_C + I_E + I_P(r, g, b) = I_C + I_E + I_P(0, 0, 0) + I_{P'}(r, g, b),$$

$$T = I_C + I_E + I_P(0, 0, 0),$$

$$I_P'(r, g, b) = I_L - T. \qquad (3)$$

where T is threshold value and $I_P'(r, g, b)$ is the color influences of the projector light.

Threshold value T can be obtained by referencing the color of the white table which is projected on black color, (0,0,0) in RGB color space.

3.4 Rotation Estimation

The combination of two colors in the markers can efficiently represent the marker's direction. We used center of weight of each color because of its rotational invariability. Initial point A is center of weight of large area. Terminal point B is center of weight of small area. If we know the coordinates of two points, direction vector can be generated from A to B. The program saves the rotation vector as a normalized unit vector form.

3.5 Marker Information Management

The marker information manager watches the detected positional data of color clusters and determines the marker ID. If the ID is equal, the manager calculates non-weighted average of positional data and sends the normalized rotation vector to the extended Kalman filter since we want visually smooth tracking data. Although the data doesn't come in time due to a slight occlusion problem, the Kalman filter can produce the smooth natural linear motion.

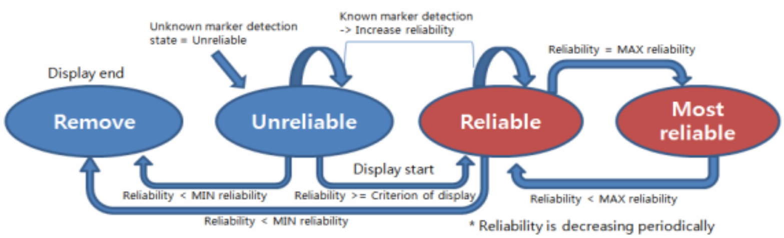


Fig. 3. Marker reliability management

The information of the marker might be occluded by the people who interact with it and by sudden changes in lighting environment. If new data comes, we generate the variable on reliability and add the marker on the marker list. We never print out the animation with new data because we don't trust the new data. If a known marker is detected, reliability of the marker will increase. The reliability of undetected marker will be decreasing periodically. When the reliability exceeds the criteria of display, the animation will be displayed, but the oldest marker does not last longer than the other markers, because of reliability of marker limits on the maximum reliability constant. Once the reliability exceeds the criteria of display, the program keeps displaying, although the reliability has fallen below the criteria of display. If the reliability has been decreased continually and become lower than the minimum reliability, the marker will be removed from the marker list. In short, if the marker satisfies the fixed criteria, the program never gives up the marker information until the marker is eliminated.

4 Implementation

The first draft of the cube was an embossed cube. The embossed cube shows the colors around the cube. We can detect the marker using relationship of the surrounded color without considering the shape of the marker and the color of the top. It is the best way for the technicians and the algorithm robustness. Artists redesigned the cube and the final draft of the cube is shown in Fig. 5. The white table was made of wood (120 x 90 x 70 cm). To soften the impact of cube throwing and to reduce influences from the specular light, we set a sheet of paper attached with rubber on the table projected by projector.

Table 1. Implementation environment

O/S	Windows 7 Ultimate K(32bit)
CPU	Intel(R) Core(TM)2 Quad CPU @2.83GHz
Main memory	DDR3 3.25GB(32bit OS)
Camera	Dragonfly Flea2G
Beam projector	Optoma DLP EP-776

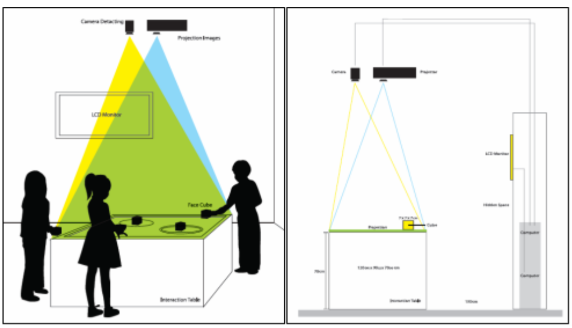


Fig. 4. Overall hardware settings

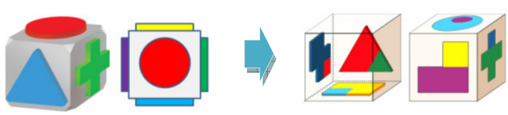


Fig. 5. First and final drafts of cube design

4.1 Animation

We designed 36 images for 24 frame/s animation for 1.5 seconds for each character. As shown in Fig. 6, we prepared 36 animations (6 animations for each side). The five themes of the animation are birds, fish, animals, boys and girls.

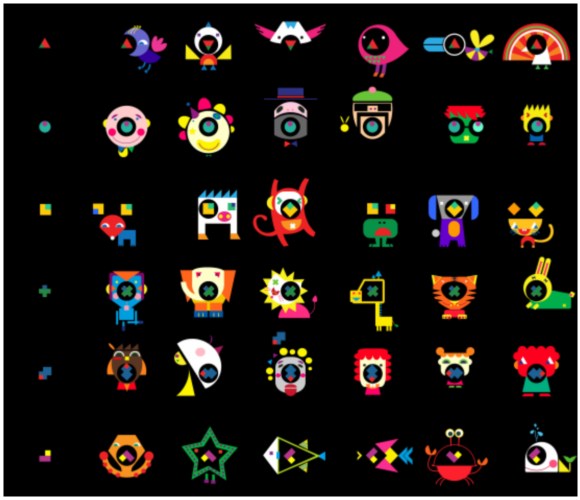


Fig. 6. 36 characters

4.2 Image Pre-processing

The program extract regions of interest, saves a new image plane, references the color of the table to remove the color influences from projector light and shadows. After thresholding, the program applies contrast stretching to thresholded image using threshold value T as a maximum intensity.

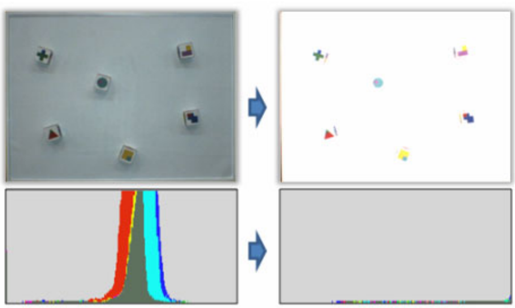


Fig. 7. Thresholding and contrast stretching result

4.3 Marker Detection

In the block-based image searching, efficiency and execution time depend on the size of the block. The distance between projection plane and camera is fixed, so we can estimate the proper size of the block and we fixed it as 26x26.

The program already knew about color distribution in each marker. It was measured by the camera beforehand. There are many errors in the cube made by manual industry, so we couldn't measure accurate area of each marker with manual measuring tools. The area of the marker is used to detect markers in camera image. If we perform the histogram matching method with 26x26 size block in 640x480 resolution image, about five billion operations will occur per 24 frames, which is prohibitive. Thus, we need algorithms to reduce the number of operations.

(1) Lowering the detecting resolution
(2) Skip counting unrelated colors during marker detection
(3) Skip detecting the block colored with achromatic at the center area of the block

Table 2. Average execution time for marker detection

Detecting resolution	Achromatic block skip	Average execution time
640x480	X	1123 ms
320x240	X	289 ms
160x120	X	66 ms
640x480	O	variable(1~1150 ms)

Supposing there are six cubes on the table, the image area of six cubes is 26*26*6 pixel, not considering noise margin. Estimating with measured data, it will take 15 ms. But, we need more to implement, because if children put their hands on the camera image to interact with cubes. The execution time will increase rapidly over 100 ms. Therefore, to guarantee maximum execution time (40 ms), we decreased detecting resolution to 160x120.

The program detects the marker on the table, and sends position and rotation data to marker information manager. The marker information manager disentangles the information from the data.

4.4 Exhibition at the Hanbit Media Gallery

The well animated characters attract visitors visually. For visitors' convenience, we posted how to interact with exhibits on the notice board.

Fig. 8. Exhibition at the Hanbit media gallery (2010.2.10-3.20)

5 Discussion

The basic color of a person's skin is mixed with red and yellow. The markers with red or yellow therefore sometimes cause misrecognition. To fix this problem, we needed to change base colors again or do a template matching method on the detected marker. However, in the template matching method, marker detection criteria and detection time will increase. They may disturb the real-time interaction with visitors. We may solve them all by GPU programming method or multi-core-based programming method, not single core solution.

The Face Cube requires a story to become a more interactive media art. We just show the projected animations to children. There is no relationship between animated characters and there is no interactivity between them too. To expand children's imagination further, we must develop much more relevant stories.

Acknowledgement

This research is supported by the Ministry of Culture, Sports and Tourism (MICST) and the Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2009 (2010).

References

1. Music Technology Group University at Pompeu Fabra. The reacTable: A Collaborative Musical Instrument. In: WETICE 2006 (2006)
2. Kakehi, Y., Iida, M., Naemura, T., Matsushita, M.: Tablescape Plus: Upstanding Tiny Displays on Tabletop Display. In: SIGGRAPH 2006 (2006)
3. Lozano-Hemmer, R.: `http://www.lozano-hemmer.com/english/projects.htm`
4. Guse, D.: Tangible table, `http://www.danielguse.de/tangibletable.php`
5. Ganser Schwab, C., Steinemann, A., Kunz, A.: InfrActables: Multi-User Tracking System for Interactive Surfaces. In: IEEE VR 2006 (2006)
6. Rick, J., Harris, A., Marshall, P., Fleck, R., Yuill, N., Rogers, Y.: Children designing together on a multi-touch tabletop: an analysis of spatial orientation and user interactions. In: Proceedings of the 8th International Conference on Interaction Design and Children 2009 (2009)
7. Pedersen, E.W., Hornbæk, K.: mixiTUI: a tangible sequencer for electronic live performances. In: Proceedings of the 3rd International Conference on Tangible and Embedded Interaction 2009 (2009)
8. Seo, H.-j.: Snow White on the Table. In: Seoul International Media Art Biennale 2008 (2008)

Entertaining Education: User Friendly Cutting Interface for Digital Textbooks

Gahee Kim, Hyun-Roc Yang, Kyung-Kyu Kang, and Dongho Kim

Soongsil University
511 Sangdo-dong Dongjak-gu, South Korea
{symple87,penril87,rcrookie,prof}@magiclab.kr

Abstract. Nowadays, the new paradigm demands digital textbooks which contain interactive contents. Our goal is to design the digital textbook providing effective multimedia and cutting interface for interactive education. To achieve this purpose, we propose the user friendly cutting interface and interactive animation for digital textbook. This interface complemented current digital textbook interface which is mostly in text based. We will discuss effectiveness of our interface for elementary students and how much our interface gives positive effect in learning.

Keywords: Digital Textbook, 3D Object Cutting, E-learning, Edutainment, Interface.

1 Introduction

Nowadays, students live in different environment from older generation. Every family has computers, and students carry a cell phone every day. They are accustomed to using computer and cell-phone, and have an ability to manage a lot of information in digital. This changed environment changed a previous teaching paradigm which is based on the off-line text book. In digital learning environment, teachers use computers and projectors and students attain information through internet and multimedia contents such as video clips, animations, and virtual reality. The paper-based textbook only delivers knowledge in one direction from teacher to students. However, the digital textbook encourages students' participation to attain the knowledge by using interactive simulation method and multimedia. In this reason, the digital textbook increases student's problem solving competency [1].

Many digital textbooks are coming out and students can easily use the digital textbooks. However, most of text books are in .pdf representation of a traditional text. These e-books may have an advantage that lowers the cost of the textbook [2], but cannot fully utilize the merit of the digital textbook. If the digital text books do not utilize multimedia to give an immersive experience, it cannot overcome the limitations of conventional textbooks.

In our research, we set our goal to develop a digital textbook which teaches 'volcano and rock' in the fifth grade curriculum. We focused on providing not only rich information based on paper-based textbook but also entertaining simulation method

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 405–412, 2010.

which can maximize users' participation and educational effect. For example, students can observe long time process of natural phenomena such as volcano creation through 3D simulation in our digital textbook. To start volcano creating simulation the user should put magma chamber under the bedrock. We put this process on purpose to increase users' participation. Our program also provides the cutting effect to improve users' concentration by entertaining the user. The user can enjoy with this cutting method by cutting the earth and volcano in various shapes. The user can also observe the cut cross section right after cutting the earth and volcano. This gives the user information about earth and volcano's inner appearance. For this reason our cutting method provides both entertaining and educational effect. At last, our digital textbook offers rich information in text and voice.

This research is executed as a portion of 'Developing essential technology for digital textbook and u-learning revitalization' under the sponsorship of the Korean government. In this paper, we will describe how we implemented an effective digital textbook interface for the fifth grade students.

2 Related Work

2.1 Digital Textbooks in Korea

KERIS(Korea Education and Research Information Service) has started providing Korean, English, Mathematics, Music and Science digital textbooks to public elementary schools in Korea since 2007. However, the majority of these textbooks illustrates main concept of a subject in digital text format not using multimedia or animation. Even though students use the digital textbook in class, they have to study in the same way as they studied using a previous textbook. Especially, the Science digital textbook has a great potential that the user can simulate dangerous natural phenomenon such as explosion of volcano which is impossible to perform in real class room, but current digital textbook also does not support these activities. In this way, we focused on developing the Science digital textbook which supports animation that shows process of volcano explosion and interactive cutting methods which can give students immersive experience to understand concept deeply.

2.2 Effective Animation Interface in Educational Programs

There were questions if the use of animation in educational programs improves student's learning or not. Media researchers have concluded whether the students learn better with animation depends on how it used [3]. In the research about how animation can be used in ways to promote learner understanding of scientific explanations, it introduces seven principles for the use of animation in multimedia instruction [4]. These principles demonstrate that a proper combination between text, narration and animation is essential interface to promote students' deep learning. We developed our animation and interactive cutting methods in accordance to these principles. We provide detail description in text and voice about the volcano, and also interactive animation which is initiated by users' participation.

3 Interface

3.1 Effective Learning Interface

3.1.1 Content Organization Based on Paper-Based Textbook Curriculum

KERIS defines the digital textbook as comprehensive digital learning material which contains management program such as simulation, multimedia, and evaluation methods based on previous paper-based textbook [5]. In the 'Revised Direction and Overview of Operation Guidelines for ICT Education Primary and Secondary School' KERIS also stressed the importance of curriculum utilizing education and examples should be closely related to curriculum [6].

In this reason we focused on organizing our content closely related to previous paper-based textbook curriculum 'volcano and rock'. We drew up the scenario flowchart such as Fig.1. We classified our contents into two categories, one with providing detail explanation to users and the other with enhancing users' experience. Detail explanation is necessary to learn characteristics about each volcano and volcanic zone. In contrast to this, to learn the process of volcano creation and eruption having an experience to observe this process is more efficient than obtaining knowledge in explanation. We provide interactive animation and cutting methods in our research to enable the user observe volcano creation and eruption. These interactive animation and cutting methods also encourage not only users' experience but also users' participation.

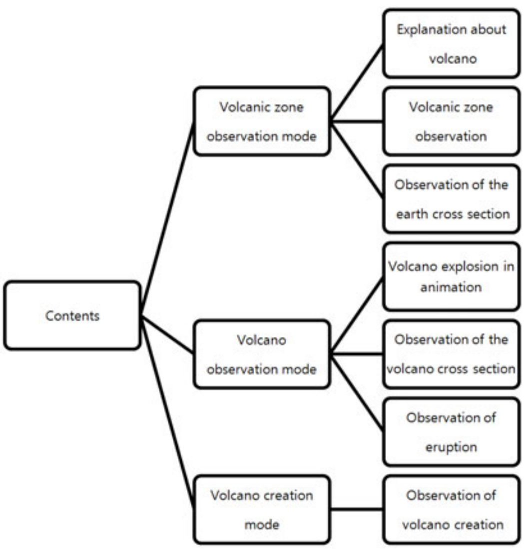


Fig. 1. Content flowchart closely related to curriculum

3.1.2 Detail Description Based on Multimedia

Our program illustrates a characteristic of the selected volcano and volcanic zone with both voice and text. In this way, students can attain more detail and accurate information about the volcano. The explanation in text and voice gives familiarity to users by enabling user to think that they are studying in previous off-line teaching method. For example, learning through the description in text is similar to study through paper-based textbooks, and hearing explanation from voice is close to learning from teachers.

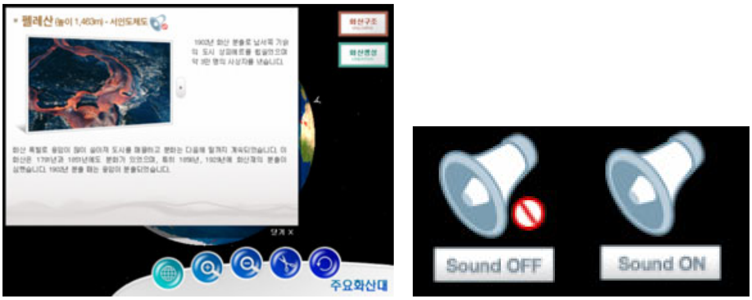


Fig. 2. Explanation on the selected volcano in text and voice

3.2 User Friendly Interface

3.2.1 Encouraging Users to Entertain Our Program: Interactive Cutting Effect

Since the users of our program are elementary students, it is very important to have an entertaining activity in our program. This fun activity will concentrate users' attention and encourage them to focus on our program for a long time. In this reason, we developed our interactive cutting method. This method not only entertains the user but also give a lesson about inside appearance of the earth and volcano. In the interactive cutting method, the user is able to cut the earth and volcano by themselves with straight and curve strokes. To use this function in our program, the user should select the cutting mode first, and cut the earth and volcano by drawing a line with a tablet pen or a mouse. The user can have fun with cutting in various shapes and enjoy observing the result of the various cut cross sections. Fig.3 shows the cut cross section of the earth and volcano which are cut by straight and curve strokes. This interesting cut cross section is generated automatically right after the user cut the earth and volcano.

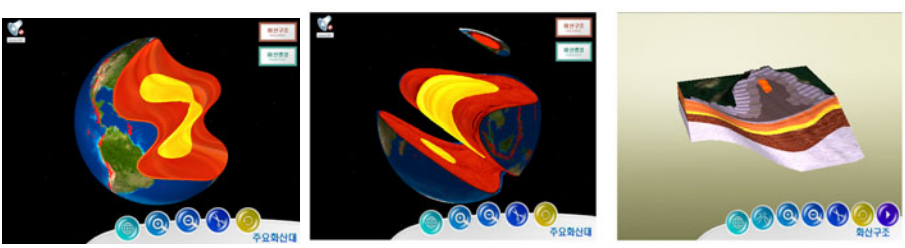


Fig. 3. Cut cross section of the earth and volcano in various shapes

3.2.2 Intuitive Manipulating Interface: Tablet Pen and Manipulating Panel

Tablet pen

We used a tablet pen as a stroke manipulating interface because this can give the user more intuitive sensing than a mouse. The user usually uses a pen when they study in a traditional way, so using the tablet pen also can give familiarity to users. The intuition and familiarity encourage the user to participate in the program more easily.

Manipulating panel

Our program also has intuitive manipulating interface such as Fig.4. Because our users are elementary students, we made easy-to-understand thumbnails as a manipulating panel. Our program provides magnification/minification, cutting and rotation methods as default methods. The user can cut the earth and volcano by pushing a sizzle thumbnail button, and observe the cut cross section closely by rotating and magnifying/minifying the cross section. When the user wants to watch simulation in animation, the user selects a play button thumbnail.

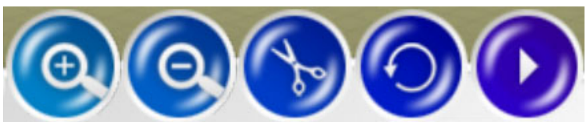


Fig. 4. Intuitive manipulating panel

3.2.3 Increasing the User Participation: Interactive Animation

When students push the button on the top right, the mode chances to the volcano observation mode where we can observe the volcano. In this mode the user can observe the process of volcano eruption and creation in animation.

The user can watch Fig.5, volcano eruption by pushing the play button in volcano observation mode.

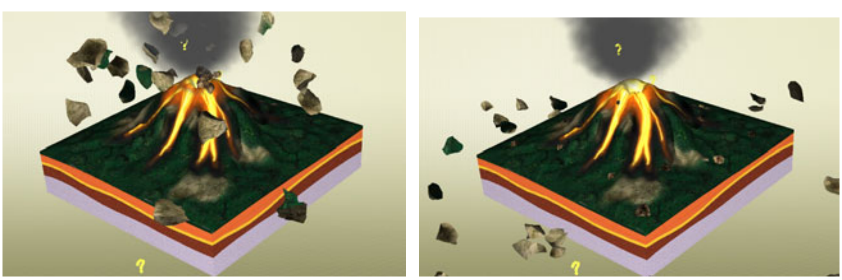
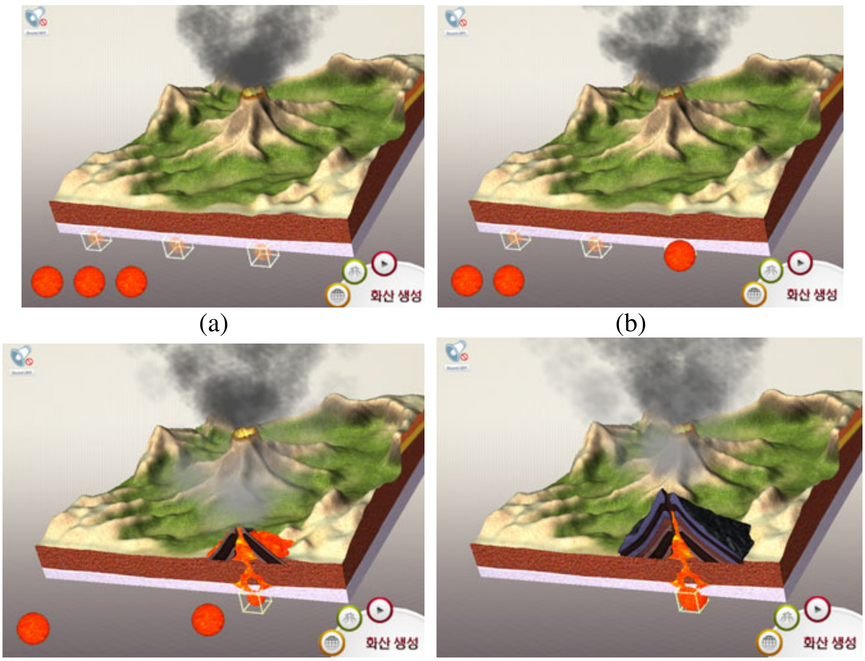


Fig. 5. Volcano eruption in animation

When the user pushes the volcano creation button, the mode is changed to the volcano creation mode. This mode enables the user to observe the process of the volcano creation. Fig.6 shows the volcano creation mode. To observe the volcano creation in animation the user should put magma chamber under the bedrock before pushing start button. This process not only encourages user's participation but also gives a lesson that magma is required to generate volcano. After setting the magma chamber under the bedrock the user pushes play button to start the volcano creating animation. If the user put one magma chamber, the process of volcano creation goes on such as Fig.6(c). The more users put magma chamber under the bedrock, the bigger volcano creates such as Fig.6(d).



(a) (b)

(c) (d)

Fig. 6. Volcano creating process in animation

4 Result

4.1 Cutting Performance Analysis

The time takes to cut the object is proportion to the number of polygons. We conducted this experiment on Intel Core2Duo 6600 CPU and 4GB Ram. The earth model consists of 14,409 polygons and the average time to cut the object is 631ms. The Volcano model consists of 29,119 polygons and the average time to cut the object is 945ms.

4.2 Trial Service Result

We conducted a trial service at one of the elementary schools in Daejeon, Korea. 26 elementary school students and 3 teachers participated in the service. After this trial service, we gave our questionnaire to students and teachers. This questionnaire is made based on the research about measuring the effectiveness of digital textbook [7]. We focused on estimating participants' educational achievement by using our digital textbook compare to previous textbook. We made 13 questions and categorized these questions into three categories such as interface, educational aspect, and satisfaction to previous textbook. The score range is set from 1 to 5. The participants gave 1 for very poor score and 5 for very satisfied score.

In the interface question we asked about the effectiveness of our cutting method, and whether it worked properly or not. We also asked if the manipulation of our digital textbook was easy to the participants. As shown in Fig.7, the participants show high satisfaction in our interface. They replied that the cutting method stimulated their curiosity, so they could participate in our program more actively. Also the thumbnail of our manipulation panel helped them to figure out the way to handle our program more intuitively.

The participants also show very satisfactory result in educational effect. In this question we concentrated on estimating the participants' understanding about 'volcano and rock' through our digital textbook. The participants replied that the animation was really effective to understand the process of volcano creation. Also they said the process that the user looks through the earth and selects the volcano they want to attain information was very helpful. This self-directed learning encourages the user to study more interactively and actively.

The average score of satisfaction compare to previous textbook is lower than other options. To complement the weak point, more interesting and entertaining aspects such as using augmented or virtual reality need to be added.

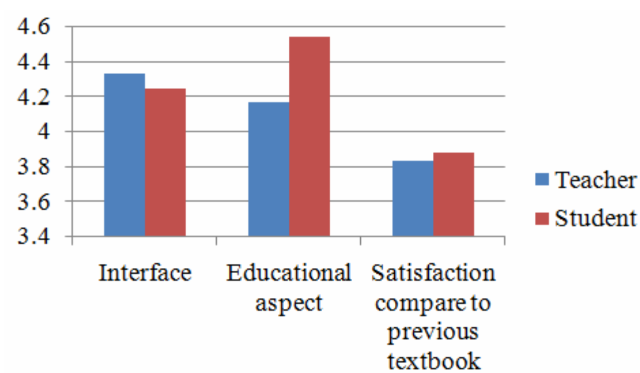


Fig. 7. Satisfaction to our system

5 Conclusion

In this paper, we introduced the user friendly cutting interface for digital textbook. This digital textbook teaches about 'volcano and rocks' content which is for fifth grade elementary school students. We attracted the users by providing cutting method and interactive animation. Our digital textbook also offers enrich information closely relate to previous paper-based textbook curriculum. This content will be serviced to the elementary students throughout the nation and the validity of application has been verified by the trial service on section 5.2. If the technology tool for encouraging interactivity such as augmented or virtual reality is added, the more effective educational content will be possible.

References

1. Song, H.D., Park, J.H.: Effects of Types of Digital Textbook Use and Levels of Academic Achievement on Learning Outcomes -from Instructional Guidance Perspective. Korean Association for Educational Information and Media 18, 29–46 (2009)
2. Kim, H.J., Seo, S.M.: A Tendency of Open Textbook in United States and Free Digital Textbook Initiative in California. KERIS Issue Report, RM 2009-36 (2009)
3. Mayer, R.E.: Multimedia learning: Are we asking the right questions? Educ. Psychol. 32, 1–19 (1997)
4. Mayer, R.E., Moreno, R.: Animation as an Aid to Multimedia Learning. Educ. Psychol. Review 14, 87–99 (2002)
5. Kim, J.E.: User Interface Design Research Based on the Digital Textbook Development. Master's thesis, Hanyang University (2008)
6. Kim, S.-M.: Revised Direction and Overview of Operation Guidelines for ICT Education in Primary and Secondary Schools. J. KERIS. Biol. 8, 6–8 (2006)
7. Yeon, K.-N.: Measuring the Effectiveness of Digital-Textbook in U-Learning Environments. Master's thesis, Donggook University (2007)

Affective Interacting Art

Youngmi Kim and Jongsoo Choi

Graduate School of Advanced Imaging Science, Multimedia, and Film,
Chung-Ang University,
221 Huksuk-dong, Dongjak-gu, 156-756, Seoul, Korea(R.O.K)
{frontier,jschoi}@imagelab.cau.ac.kr

Abstract. This paper to study the potential of expressing ink-and-wash painting through interaction, and present a direction that can coincide with modern paintings by developing ink-and-wash painting from a traditional aspect through analyzing the theories and tech-niques instilled in my works.This work is an interactive visualization of an oriental cymbidium using modern technology which our oriental ancestors painted for mental training. During the old days in the orient, people used to wipe cymbidium leaves or painted cymbidium for mental training by having a cymbidium always by their side. Through the act of wiping cymbidium leaves with utmost care, a cymbidium instilled with ancient philosophical ideas is visualized.

Keywords: Ethnographic Design, mental training, Interactive art, oriental digital art, ink-and-wash painting, haptic.

1 Introduction

In the orient, calligraphy is utilized for the mastering of ones mind rather than just simply writing a letter or drawing a painting well. One must focus and possess a proper attitude when engaging in calligraphy because process is given more emphasis over results. This principle applies in drawing the cymbidium leaves where the participants mentality will be projected on the screen. Moreover, in order to complete a balanced cymbidium painting, cymbidium leaves must be evenly selected, and the direction and angle one wants to draw must be adjusted. This is easily linked to emotions, desires, and moods, with regenerative senses on exercise as its medium. Also, almost an unlimited expression of ones mind can be possibly conveyed by using different techniques such as direction, speed, power, and rhythm. In particular, emotional and physical safety can be obtained through aesthetic perspective, fine motor skills, hand-eye coordination, and sense of achievement. In this way, art with oriental features can be enough driving force to arouse interest and curiosity to national culture of children who will be responsible for future generations. This may be achieved by active utilization of media in line with modern perspectives.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 413–415, 2010.

2 Meaning of the Act of Drawing a Cymbidium

The spirituality, idea, and unique characteristic of our ink-and-wash painting will be examined. This means that I intend to find the characteristic of ink-and-wash painting from spirituality and express the spirituality through ink-and-wash painting in my work in order to shed new light on the features of ink-and-wash painting from a modern point of view. Prior to this, I would like to help understand the act of wiping cymbidium as a means of interaction. Asians has been living with cymbidium as if it is a close friend, and an oriental cymbidium is always placed in noble places. The act of wiping each cymbidium leaves with human hands as if sweeping it off is not to maintain cleanliness but rather to uphold the spirit instilled within the cymbidium. In other words, if one's heart is troubled and there are a lot on ones mind, mental training was carried out with a careful act of wiping cymbidium leaves with a calm mind by having a cymbidium close to oneself. Once each cymbidium is wiped off, the things troubling ones mind is forgotten and ones heart is emptied. This act has great meaning as it plays a role of interacting with the cymbidium painting. While the meaning of a cymbidium painting is also similar to the act of wiping a cymbidium, it is instilled with more meaning. The artist always conducts close observation of the subject before painting it. However, the oriental method of sketching is not depicting the subject while looking at it in detail but making a sketch by memory apart from the subject. Oriental drawing technique expands the artists creative space while relying on visual memory and shape memory by making the most use of ones subjectivity and imagination breaking away from the restraint of the actual subject. What is important is drawing with creativity yet keeping several rules. The curves, and bold and thick lines of a cymbidium must be drawn, and its leaves must shake in the wind and have bones. In this work, cymbidium leaves are visualized while being divided up into leaves curved according to the angle of wiping up or down and cymbidium leaves blowing in the wind. Just like this, even in the writers work, a poem and a writing about its impression going well with the style of painting after the cymbidium and flower is painted is written in harmony and last of all, a piece of an artwork is finished by stamping a seal.

3 Design

When a cymbidium leaf is selected and wiped using the thumb and index finger as if sweeping it down, a cymbidium leaf is drawn one after another inside the digital frame. A flexible sensor is attached behind cymbidium leaves, and the size of an angle made while a cymbidium leaf is curved appears due to the difference in electrical current through flexible sensors. The micro controller detects this signal and converts the analog signal to a digital signal. Then, the data is sent to the PC using UART communication. The server program can detect UART output signal sent to the desktop from the micro controller and converts this data to a TCP/IP data format. As a result, this format can send this data set from the server to the client. Consequently, the client draws the painting according to this signal after detecting the signal.

When the angle of the curve communicated through the cymbidium leaf is big, the cymbidium leaf in the painting is drawn as curved leaves or leaves blowing in the wind. A total of 8 cymbidium leaves are drawn up, and if one breathes air into a flower after drawing a cymbidium leaf, a floral axis and flower is drawn. It takes in the strength of the breath and when it is strong, a full blossomed flower is visualized and when the breath is weak, an unblossomed flower is visualized. When the interactive work is finished, a caption or a poem that goes well with the style of the painting is written in the remaining blank space with a stamped red seal. All of these methods communicate that the painting was produced in the same way cymbidium paintings were painted in the old days. The regularity and creativity of the cymbidium painting striking a balance, writing a poem and writing which goes well with the blank space, the seal of the artist, and above all, the greatest significance is that interaction was achieved through the act of wiping the cymbidium leaves with the attitude and spirituality of the person drawing the painting.

4 Conclusions

This is an interactive well-being design made so children can engage in mental training physically and mentally through this artwork and naturally have an interest in the Ethnographic aspect instilled in the work. In this artwork, Education is carried out through a series of process from free selection, activity, repetition, habit to learning. This process provides special experience unobtainable in daily life, and will help setting up an educational set of values in establishing an attitude of learning during adolescence later on. To children, play doesnt simply mean playing, but means they learn sociality, emotions, reasoning, and personal relationships. Furthermore, children expressing their experience and emotions with play itself is a process of self-healing. Feeling the joy of learning, a child freely engaging in work one desires and doing an assignment with their own efforts helps display each childs intellectual ability and mentality in their individual stage of development. Therefore, it will contribute to enhancing the development of a childs emotional quotient (E.Q.).

Acknowledgments. This work was supported by Korean Research Foundation under BK21 project, Seoul R&BD Program(TR080601), Seoul Future Contents Convergence (SFCC) Cluster established by Seoul R&BD Program(10570).

Flexible Harmonic Temporal Structure for Modeling Musical Instrument

Jun Wu, Yu Kitano, Takuya Nishimoto, Nobutaka Ono, and Shigeki Sagayama

The Graduate School of Information Science and Technology, University of Tokyo,
Tokyo 113–8656, Japan
{wu,kitano,nishi,onono,sagayama}@hil.t.u-tokyo.ac.jp

Abstract. Multipitch estimation is an important and difficult problem in entertainment computing. In this paper a flexible harmonic temporal structure for modeling musical instrument was proposed for estimating pitch in real music. Unlike the previous research, the proposed model does multipitch estimation according to the specific characteristics of specific musical instrument and uses EM algorithm to estimate the parameters in the model. Through choosing parameters suitable for its own characters for specific instrument, the proposed model preponderated over the common model.

1 Introduction

Multipitch estimation is one of the fast growing topic in MIR in which most of the research going on now is using signal processing method. A multipitch tracking method in noisy environment by filter bank process and pitch tracking using HMM was proposed by Wu et al. [1]. Goto presented a method to track F0 of objective single sound from polyphonic musical signals without restriction of the number of simultaneous sounds [2]. Some other multipitch analyzers such as graphical model-based [3], filterbank-based [4], nonparametric Kalman filtering-based [5], [6]. Then a method for multipitch analysis called Harmonic-Temporal Clustering (HTC) was proposed [7] to deal with the harmonic and temporal structures in both time and frequency directions and shows high performance. However, all of these algorithms cannot do multipitch estimation according to the specific characteristics of specific musical instrument, which is actually important and meaningful for real music. In this paper, a flexible harmonic temporal structure for modeling musical instrument is proposed in section 2. The proposed model is based on the clustering principle and uses EM algorithm to estimate each mean parameter. In section 3, the experimental results were demonstrated.

2 Flexible Harmonic Temporal Structure

We propose model $q_k(x, t; \theta)$ to be the model for a single note in the music, where x is log-frequency, t is time and θ is the parameter in the model. It is composed of the fundamental partial and harmonic partials. The normalized energy density of the nth partial in the kth source model can be assumed to be a multiplication of the power envelope of the nth partial $U_{k,n}(t)$ and the Gaussian distribution centered at

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 416–418, 2010.

$$\mu_k\,(\mathrm{t}) + \log(n),\ U_{k,n}(t) \times \frac{v_{k,n}}{\sqrt{2\pi}\sigma_k} e^{-(x-\mu_k(t)-logn)^2/2\sigma_k^2} \quad n = 1, \dots, N \tag{1}$$

$\mu_k\,(\mathrm{t})$ is pitch contour of the *k*th source, $v_{k,n}$ is relative energy of *n*th partial in *k*th source. Let the frequency spread of each harmonic component be approximated by a Gaussian distribution function when the spectra are obtained by the wavelet transform (constant Q transform) using Gabor wavelet basis function. Denote $U_{k,n}(t)$ as the power envelope of the *n*th partial.

$$U_{k,n}(t) = \sum\nolimits_{\forall y} \frac{u_{k,n,y}}{\sqrt{2\pi\phi_{k,n}{}^2}} exp\left\{-\frac{(t-\tau_k-y\phi_{k,n,y})^2}{2\phi_{k,n}{}^2}\right\} \tag{2}$$

τ_k is the center of the forefront Gaussian, which is considered as an onset time estimate, $u_{k,n,y}$ is the weight parameter for each kernel, which allows the function to have variable shapes for each harmonic partial.

$u_{k,n,y}$ should be normalized to satisfy $\forall k, \forall y\colon \sum_y u_{k,n,y}\,(x,t) = 1$. $\phi_{k,n,y}$ is the distance between the centers of the Gaussian function kernels. The power spectrogram structures of different instruments are very different. So we set $\phi_{k,n,y-1} = \alpha\phi_{k,n,y-1}$ to give flexibility to the distance between the Gaussian function kernels. If the instrument's power spectrogram structure is relatively steep at the beginning part, the model is able to choose larger α which means more Gaussian function kernels at the steep beginning part.

The source models $q_k(x,t;\theta)$ are expressed as a mixture of Gaussian mixture model (GMM) with constraints on the kernel distributions: supposing that there is harmonicity with *N* partials modeled in the frequency direction, and the power envelope is described using *Y* kernel distribution in the time direction. The source model can be written in the form

$$q_k(x,t;\theta) = \sum\nolimits_n \sum\nolimits_y S_{k,n,y}(x,t;\theta) \tag{3}$$

And the Kernel distribution can be written in the form

$$S_{k,n,y}(x,t;\theta) = \frac{w_k v_{k,n} u_{k,n,y}}{2\pi\delta_k \emptyset_k} e^{-\frac{(x-\mu_k(t)-logn)^2}{2\sigma_k^2} - \frac{(t-\tau_k-y\emptyset_{k,n})^2}{2\phi_{k,n,y}^2}} \tag{4}$$

w_k is the energy of the *k*th source. Therefore the source model $q_k(x,t;\theta)$ is the mixture of mixture of Gaussian distribution $S_{k,n,y}(x,t;\theta)$. And the whole model is the mixture of the source model $q_k(x,t;\theta)$.

The proposed the algorithm uses EM procedure for the parameter estimation procedure. We assume that the energy density *W(x;t)* has an unknown fuzzy membership to the *k*th source, introduced as a spectral masking function $m_k(x,t)$. To minimize the difference between the observed power spectrogram time series *W(x;t)* and the model $\sum_k q_k(x,t;\theta)$, we use the Kullback–Leibler (KL) divergence as the global cost function.

$$J \sum\nolimits_k \iint_D m_k(x,t) W(x;t) log \frac{m_k(x,t)W(x;t)}{q_k(x,t;\theta)} \tag{5}$$

Satisfying with:

$$\forall x, \forall t, \sum\nolimits_k m_k\,(x,t) = 1, 0 < m_k\,(x,t) < 1.$$

Then the problem is regarded as the minimization of (5).

The membership degree $m_k(x,t)$(spectral masking function) of kth source/stream can be considered to be the weight of the kth source model in the whole spectrogram model. It is unknown at the begging and need to be estimated. On the other hand, the spectrogram of the kth source can be modeled by a function $q_k(x,t;\theta)$, where θ is the set of model parameters. They are also unknown variables. The proposed model works by using EM algorithm for iteratively updating of: E-step: $m_k(x,t)$ with θ fixed and M-step: θ with $m_k(x,t)$ fixed. $m_{k,n,y}(x,t)$ is masking function.

The E-step is realized by the following equation.

$$m_k(x,t)m_{k,n,y}(x,t) = \frac{s_{k,n,y}(x,t;\theta)}{\sum_k \sum_n \sum_y s_{k,n,y}(x,t;\theta)} \quad (6)$$

The update of parameters can be obtained analytically by undetermined multipliers Lagrange's method.

3 Experiments

To evaluate the proposed flexible harmonic temporal structure, we tested it with 151 music instrument pieces (including 4 instruments: 36 guitar pieces, 45 violin pieces, 36 flute pieces and 34 oboe pieces) chosen from the RWC music database. [8] Since the RWC database also includes the MIDI files associated with each real-performed music signal data, we evaluated the accuracy by comparing the estimated fundamental frequency and the MIDI files. The proposed algorithm preponderates over the HTC [7] approach for 23.3% for guitar signal, 2.8 % for oboe signal, 12.5 % for flute signal and 3.1% for violin signal.

References

1. Wu, M., Wang, D., Brown, G.J.: A Multi-pitch Tracking Algorithm for Noisy Speech. In: ICASSP 2002, vol. 1, pp. 369–372 (1995)
2. Goto, M.: A Predominant-F0 Estimation Method for CD Recordings: MAP Estimation Using EM Algorithm for Adaptive Tone Models. In: Proc. ICASSP 2001, vol. 5, pp. 3365–3368 (September 2001)
3. Kashino, K., Nakadai, K., Tanaka, H.: Organization of hierarchical perceptual sounds: Music scene analysis with autonomous processing modules and a quantitive information integration mechanism. In: Proc. IJCAI, vol. 1, pp. 158–164 (1995)
4. Klapuri, A., Virtanen, T., Holm, J.: Robust multipitch estimation for the analysis and manipulation of polyphonic musical signals. In: Proc. COST-G6 Conf. Digital Audio Effects, pp. 233–236 (2000)
5. Nishi, K., Ando, S.: Optimum harmonics tracking filter for auditory scene analysis. In: Proc. IEEE ICASSP 1996, vol. 1, pp. 573–576 (1996)
6. Abe, M., Ando, S.: Auditory scene analysis based on time-frequency integration of shared FM and AM (II): Optimum time-domain integration and stream sound reconstruction. Trans. IEICE J83-D-II(2), 468–477 (2000) (in Japanese)
7. Kameoka, H., Nishimoto, T., Sagayama, S.: A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering. IEEE Trans. on Audio, Speech and Language Processing 15(3), 982–994 (2007)
8. Goto, M., Hashiguchi, H., Nishimura, T., Oka, R.: RWC music database: Popular, classical, and jazz music database. In: Proc. ISMIR, pp. 287–288 (2002)

Towards a Service-Oriented Architecture for Interactive Ubiquitous Entertainment Systems

Giovanni Cagalaban and Seoksoo Kim*

Department of Multimedia, Hannam University, Daejeon, Korea
gcagalaban@yahoo.com, sskim0123@naver.com

Abstract. Ubiquitous computing is not only applied to doing daily activities and integrated into everyday objects but for entertainment, and gaming as well. In this research, we explore the relevance of ubiquitous computing to entertainment systems using devices such as mobile devices. We introduce a service-oriented architecture for ubiquitous entertainment systems to establish collaborative relationships between heterogeneous devices to provide users an interactive and ubiquitous entertainment and fun.

Keywords: ubiquitous entertainment, service-oriented architecture, mobile device.

1 Introduction

The vision of having computing services readily available is becoming more of a reality [1][2]. Ubiquitous computing has also applied in ubiquitous entertainment systems [3]. The emergence of mobile devices and the recent trend in 3G technology enables people to work, play games, or entertain themselves at home, office, or even on the move [4]. The capability of wireless devices to connect to a combination of networks such as wireless fidelity (WiFi), worldwide interoperability for microwave access (WiMAX), global system for mobile communication (GSM), and Bluetooth connections allows them to interoperate in a heterogeneous environment.

In this research, we design a service-oriented architecture for ubiquitous entertainment systems to establish collaborative relationships between heterogeneous devices. The proposed service oriented architecture enables mobile users to find themselves in environments rich in ubiquitous services that their mobile devices can take advantage of for entertainment, gaming, and fun.

2 Ubiquitous Entertainment Services

The ubiquitous service center has more computing power compared to the mobile devices that people use. Examples of such ubiquitous environment may be in a conference room, coffee shop, restaurant, train station, airport, or shopping mall, etc. The ubiquitous service centers can be configured in controlled and relatively stable environments, such as homes, offices, or public areas that experience a high density

* Corresponding author.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 419–421, 2010.

of mobile users to provide ubiquitous services such as shown in Fig. 1 where a user is playing a game in a mobile device surrounded by sensor network and is equipped with situation awareness to seamlessly perform other activities [5].

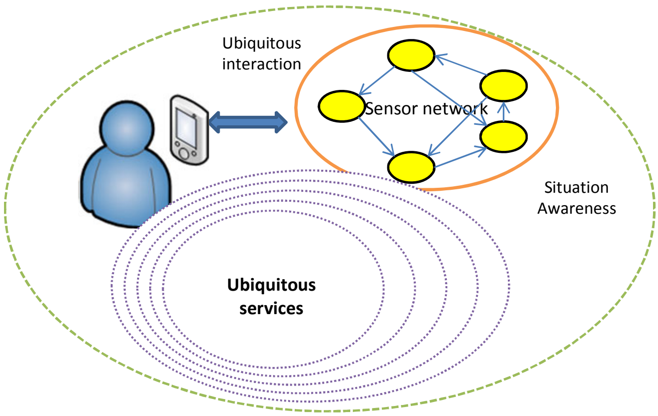


Fig. 1. Ubiquitous environment

In the ubiquitous entertainment system, a mobile user should be able to access and consume the services provided by the ubiquitous entertainment server through the ubiquitous environment in a seamless manner. The proposed service oriented architecture for the ubiquitous entertainment system is shown in Fig. 2.

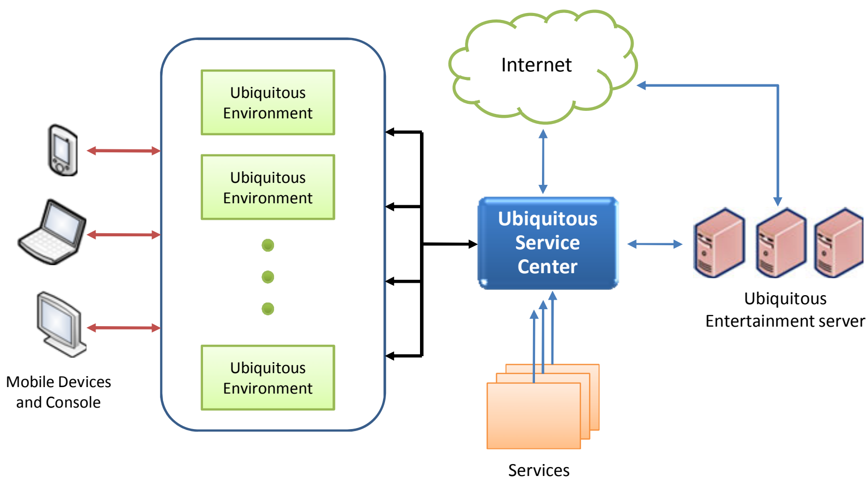


Fig. 2. Service oriented architecture for ubiquitous entertainment system

One of the functions of the ubiquitous service center is to register and query of services. The registration and discovery of services in a ubiquitous service center is done by the service manager. The service manager is responsible for facilitating the registration service where mobile devices can discover and select from a set of services. Also, ubiquitous service center is responsible for session management. If a user of a mobile device discovers that there are available services in the ubiquitous service center that are of his interest, he can query and request the service manager for services to be used by his mobile device.

In the service oriented architecture of the ubiquitous entertainment system, the main focus of the design is the service interface. We consider that all the service components are deployed in ubiquitous environment and are visible for mobile devices to invoke or consume over the network. Also, the nature of SOA is focused on creating services using components with well defined interfaces, which allows devices to be loosely coupled with each other.

3 Conclusion

We have presented the significance of introducing a service-oriented architecture for ubiquitous entertainment system. The main focus of the service oriented architecture is the creation of flexible and extensible interfaces for the interactive delivery of services for a ubiquitous entertainment. The design of a service-oriented architecture is an integral part of ubiquitous technology. In the future, we aim to improve the security design of the ubiquitous entertainment system by implementing security algorithms to enhance reliability and improved performance of ubiquitous services.

Acknowledgements

This paper has been supported by the 2010 Hannam University Research Fund.

References

1. Weiser, M.: The computer for the twenty-first century. Scientific American 265(3), 94–104 (1991)
2. Davies, N., Gellersen, H.W.: Beyond prototypes: Challenges in deploying ubiquitous systems. IEEE Pervasive Computing 1, 26–35 (2002)
3. Endres, C., Butz, A., MacWilliams, A.: A Survey of Software Infrastructures and Frameworks for Ubiquitous Computing. Mobile Information Systems Journal 1(1), 41–80 (2005)
4. Cheok, A.D., Fong, S.W., Goh, K.H., Yang, X., Liu, W., Farzbiz, F.: Human Pacman: A Sensing-based Mobile Entertainment System with Ubiquitous Computing and Tangible Interaction. In: Proceedings of the Second Workshop on Network and System Support for Games (2003)
5. Sung, J., Lopez, T.S., Kim, D.: The EPC sensor network for RFID and WSN integration infrastructure. In: IEEE International Conference on Pervasive Computing and Communications Workshops, Los Alamitos, USA, pp. 618–621 (2007)

Narrative Entertainment System with Tabletop Interface

Takashi Mori, Katsutoki Hamana, Chencheng Feng, and Jun'ichi Hoshino

1-1-#3M309, Tennodai 1-chome, Tsukuba-shi, Ibaraki, Japan
e0611423@edu.esys.tsukuba.ac.jp,
{hamana,feng_chencheng}@entcomp.esys.tsukuba.ac.jp,
jhoshino@esys.tsukuba.ac.jp

Abstract. We propose the Narrative Entertainment System with Tabletop Interface. This system is using a miniature-shaped interface called Physical Character. By recognizing the operation of Physical Character, virtual actor's behavior is under control, and that offers a method to create a story while playing with miniatures in a familiar way since our childhood. In the form of intuitive operation with this interface, you are not only an observer, but also a creator.

Keywords: narrative entertainment, tabletop, interface, tangible, interactive storytelling, edutainment.

1 Introduction

Storytelling is typical of the narrative entertainment, and is helpful to children in improving their logic, language skill, expressions and communication skills [1].

Storytelling with some interactions gives a chance to enable users to participate in the creation of the story, that has expanded the interests of story experience and reinforced the newness of the story contents. StoryMat[2] uses the toys, which can record children's voice and action for saving and replaying stories when they are playing with. On the other hand, there are researches of interactive systems by operating objects on the table like [3] and [4].

In this paper, we propose a method of Narrative Entertainment System with Tabletop Interface with a miniature-shaped interface in the real world to control the stories and environments of virtual world. In this system, miniatures and dolls with a full length of several centimeters take important part in interface to enable an interaction between users and the story world. Dolls or miniatures are familiar since our childhood, so it is an advantage to make it more intuitive than regular input devices.

As a system to control the process of the story, we have developed a story engine that can manage a huge number of episode segments, and it can actively generate animations of massive character.

2 System Concept

The picture in the leftside of Fig. 1 is showing the concept of this system. The story world consists of the part of virtual world and the part of real world. This system has

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 422–424, 2010.

an engine to manage the story (Story Engine), several miniatures to control the items or virtual actors (Physical Character), a table-shaped device with a display to get the input data (Tabletop Interface; TTI), and a vertical theater view (Story Monitor). This system is supposed to be placed in a living room. Whenever users watch the animation of life-like characters, they can move the Physical Characters to create their stories actively. By using substantial tangible interfaces such as dolls or miniatures, familiar and intuitive operation becomes possible, and this system is available to encourage children to study. Moreover, it supports a multiple play, which is related to improve children's communication skills.

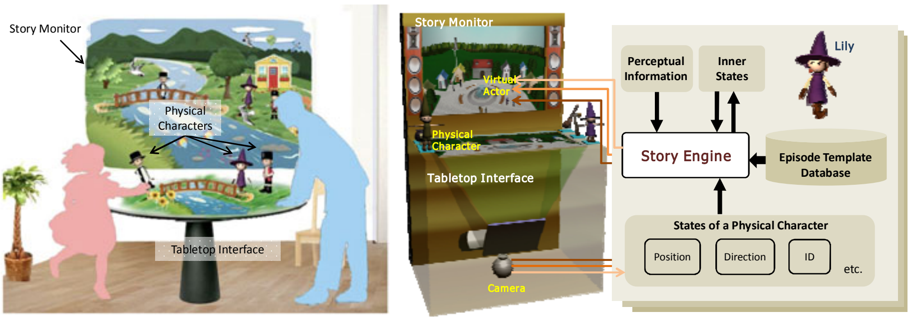


Fig. 1. System Concept Image and System Configuration

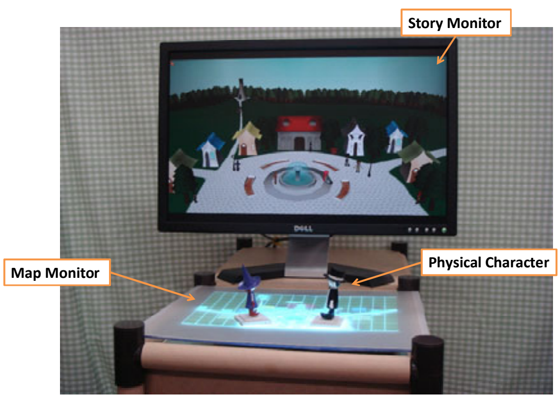


Fig. 2. TTI System Prototype

3 System Configuration

The picture in the rightside of Fig. 1 is showing composition of this system. It is mainly consist of 3 parts, an engine to manage the story, a table-shaped device with a display to get the input data, and a vertical display. Each of them are named Story Engine, Tabletop Interface (TTI), and Story Monitor. Moreover, models of virtual characters and virtual items have been used as input device. That is called Physical Character.

Story Engine controls the Character Status and Story Progression on the basis of the input data of TTI, it shows the status of characters and objects on the Story Monitor. The core system, which controls the whole action of characters to be shown in the story, includes our privious research Massive Action Control System (MACS) [5].

TTI enables a way of creating stories by manipulating Physical Characters. TTI includes Physical Characters, camera, projector and mirror. Position of Physical Characters is recognized with camera and data are sent to Story Engine and reactions are generated to respond user's operation. Fig. 2 shows the system prototype of TTI.

Story Monitor provides a real time view of the virtual world which the users have created. It mainly shows the 3D models of the virtual world, of which corresponded with Physical Character, and other characters living in the story world. Story gets the user's operation and continues the story, which is projected to the Story Monitor to help it easier to understand.

4 Conclusion

We proposed a Narrative Entertainment System with Tabletop Interface which can be intuitively operated by children and adults in this paper. By using a miniature-shaped interface, stories can be created while playing with miniatures in a familiar way since our childhood. We have constructed the prototype and tested the system with episode templates and verified the effectiveness. As the result, users evaluated the intuitive way of control and easiness of operation about this interface. However they pointed out the lack of creation of stories.

There are two directions for future development. The first is to develop episode templates. Cooperation and teachings depend on contents, so we are thinking of bring in long stories, or using episodes with clear logic to upgrade and test the episode templates. The second one is to develop the interface. The current TTI is operating in a 2D panel. We are thinking of moving and changing directions miniatures like in the entertainment of Dollhouse, which can support a 3D movement and transformation. It may be updated with 3D sensing and movable Physical Characters.

References

1. Wright, A.: Creating stories with children. Oxford University Press, Oxford (1997)
2. Ryokai, K., Cassel, J.: StoryMat: A play space with narrative memories. In: Proc. of CHI 1999, Pittsburgh, PA, pp. 272–273 (1999)
3. Kakehi, Y., Yamaoka, J., Akatsuka, D., Naemura, K.: Tablescape Animation: A Basic Study on Supporting Making Animations Using Tabletop Objects. In: Interaction 2009 (2009)
4. Kakehi, Y., Iida, M., Naemura, T., Shirai, Y., Matsushita, M., Ohguro, T.: Lumisight Table: An Interactive View-Dependent Display-Table for Mutiple Users. IEEE Computer Graphics & Applications 25(1), 48–53 (2005)
5. Hoshino, J., Hamana, K., Shiratori, K., Nakano, A.: Distributed Episode Control System for Interactive Narrative Entertainment. In: Natkin, S., Dupire, J. (eds.) ICEC 2009. LNCS, vol. 5709, pp. 145–156. Springer, Heidelberg (2009)

Automated Composing System for Sub-melody Using HMM: A Support System for Composing Music

Ryosuke Yamanishi, Keisuke Akita, and Shohei Kato

Dept. of Computer Science and Engineering, Graduate School of Engineering,
Nagoya Institute of Technology,
Gokiso-cho Showa-ku, Nagoya, 466-8555 Japan
{ryama,shohey}@juno.ics.nitech.ac.jp

Abstract. We propose an automated composing system for sub-melodies focusing especially on pitch and rhythm. We constructed the system using a Hidden Markov Model (HMM). In a composing experiment, we obtained various melodies depending on the song set used for learning, and the results suggest that this system can learn the features of song sets that are selected while considering music genres, music culture, or nuances of composers.

Keywords: automated composing, HMM, machine learning, music.

1 Introduction

Nowadays, many forms of entertainment exists, e.g., computer games, movies, and sports. Music is another form, and it has been especially loved by people all around the world since ancient times. Composing music can also be thought of as a form of entertainment as interesting as listening to music, and therefore, many people who love music have also tried to compose it. The question we have to ask here is whether people who do not have much music experience can compose songs. Most music is composed by taking many elements into consideration, e.g., rhythm, chord progression, and melody [1]. In particular, composing music that has many parts requires a high level of knowledge because there may be several different sub-melodies. However, it is difficult for people who do not have a musical background to obtain this knowledge. Some simple melodies or chord progressions can be hummed or imagined without this knowledge. We expect that music which is not by book reflects the newest styles of music. Thus, some support systems for composing music are necessary, for example the one described in reference [2].

In this paper, as the first step in composing music responding to a demand, we propose a support system for composing a sub-melody. This system focuses on pitch and rhythm, which are thought to be the most important factors in a melody, and learns them using Hidden Markov Model which is a probabilistic learning model.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 425–427, 2010.

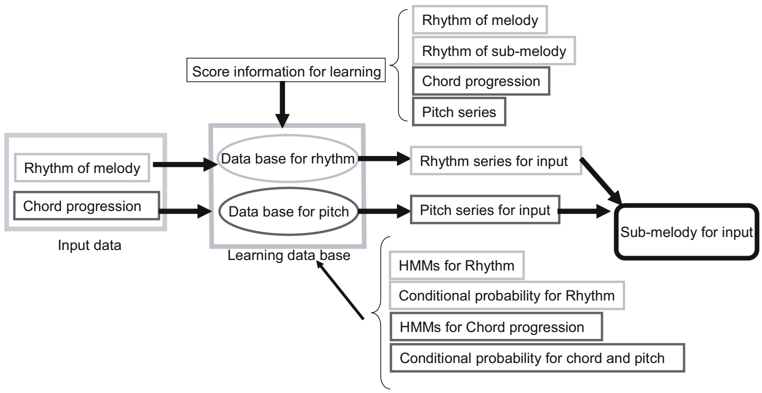


Fig. 1. Flowchart of the system

2 System

In this study, rhythm, chords, and pitch were each modeled and learnt from songs that had many parts, melodies and sub-melodies. SN indicates the song set, and score information about a song $sn \in SN$ was modeled and learnt according to the following procedure. The songs in SN were without modulation.

Rhythm series of melody and rhythm series of sub-melody are each modeled using following states, "sound", "silent", or "continuance of the previous state." In this study, we used an eighth note as the minimum unit, or beat in scores. Taken in the light of tonality, we use degrees for tonic in each bar, song sn to represent the chord's name for learning chord progressions. In this study, any bar in sn have only one chord. And then, we use three chords over the three bars as a chord progression. A pitch series of the sub-melody in each bar, song sn is learnt as a series of degree for a tonic of the song sn. Conditional probability that show the relationships between rhythm series of melody and rhythm series of sub-melody, and conditional probability that show the relationships between chord progressions and pitch series are each calculated. Additionally, score information about each bar in sn were learnt respectively using Left-to-Right HMM where the state change in a single direction.

When we input a melody and chord progression, then the system automatically outputs a sub-melody using the HMMs. Fig. 1 shows the flowchart of this system.

3 Composing Experiment

We prepared five songs from the "RWC Music Database: Jazz Music Database [3]" as the song set for jazz music, five songs from Japanese popular music randomly selected as the song set for J-POP, and five compositions by Tchaikovsky as the song set for classical. All songs used in learning had four-four time. Namely, we

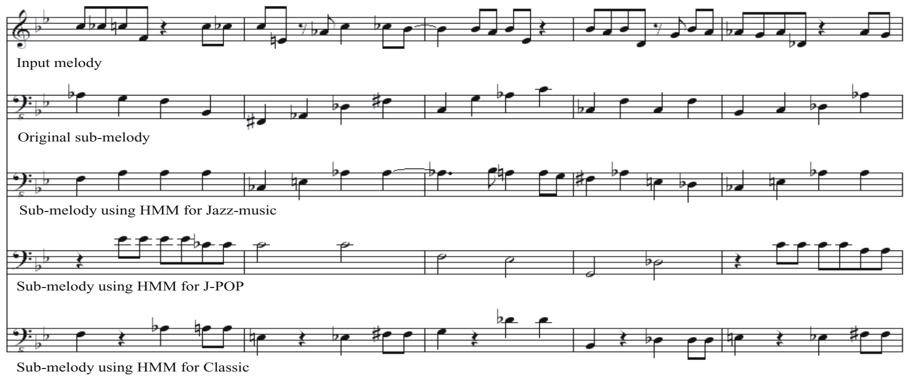



Fig. 2. Results of composing sub-melodies using various HMMs

prepared three categories of song sets to learn. In this experiment, we composed three sub-melodies with same input melodies using an HMM for jazz music, an HMM for J-POP, and an HMM for classical. We prepared a song from the "RWC Music Database: Jazz Music Database" that was not used in learning as an input melody. Part of the input melody, the original sub-melody (for reference), and the sub-melodies composed by the proposed system using each HMM are shown in Fig. 2. Although the input melodies are the same, the sub-melodies composed by using the proposed system are different due to using different HMMs.

4 Conclusion

In conclusion, we should note that the proposed system can be used as a system to support composing. In the future, we plan to expand this study and consider accessory notes while dealing with notes below the eighth notes. We will also construct a system to allow feedback between the systems and the composer when transcribing the music.

References

1. Merwe, P.V.: Origins of the Popular Style: The Antecedents of Twentieth-Century Popular Music. Oxford University Press, Oxford (1992)
2. Hasui, H., Ogura, H.: Application of Interactive Composing System, i-Sonneteer, to Other Musical Styles. Information Processing Society of Japan 2008(12), 157–162 (2008)
3. Gotou, M., Hashiguchi, H., Nishimura, T., Oka, R.: RWC Music Database: Music Genre Database and Musical Instrument Sound Database. In: Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003), pp. 229–230 (2003)

A Study on the Development of Mobile Based SNS-UCC Writing Tool

Hae Sun No[1] and Dae Woong Rhee[2]

[1] Ph.D. Course, Dept of Game, Graduate School,
SangMyung University, 7 HongJi-Dong Jongno-Gu, Seoul, Korea
hs4k2@naver.com
[2] Division of Digital Media, College of Computer Software & Digital Media,
SangMyung University 7 HongJi-Dong Jongno-Gu, Seoul, Korea
rhee219@smu.ac.kr

Abstract. Driven by the development of wireless services with the advent of smart phone along with iPhone, wireless services are drawing attentions rather than PC internet based wired services. Since this technique is a mobile device judged as being in an environment suitable to provide SNS, there also need be researches. This research proposed mobile based SNS-UCC design through mobile based SNS-UCC writing tool development cases.

Keywords: SNS, UCC, Mobile, Writing Tool, Development.

1 Introduction

Presently throughout the world, interactive services that offer amusements to users and are represented by UCC and Web 2.0, the SNS based UCC services, are extensively developed. The background SNS and UCC grew thus active is due to the desires of users to produce contents personally and consume them. As they prefer information in forms that they can understand intuitively, the types of contents change from image to moving pictures.[1] Another reason SNS and UCC grew active is that the opportunities to use wireless network increased owing to supply of 3G phone and high performance terminal devices like smart phone etc showed up.[2]

2 Development of Mobile Based SNS-UCC Writing Tool

SNS based UCC services are produced based on PC internet. However PC internet based SNS-UCC service has the following disadvantages:

The first is the inconvenience of saving personal information. In case of PC internet base, users are faced with the inconvenience that they have to sign up individually at each and every site that they intend to use, and enter and save information on themselves each time.

The second is time and space limitation. PC internet based SNS requires a space that offers PC and internet. In other words, if PC is not available, or if internet cannot be accessed though PC is available, SNS cannot be used.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 428–430, 2010.

The third is the many time and efforts needed in learning the using method. PC internet based UCC writing tool requires users to personally edit story structure or screen setting as well as moving picture, photo, speech and various effects. Therefore, many hours and efforts are required in learning and using such functions.

In developing mobile based SNS-UCC, the following designing principles should assumed as fundamental:

First, easy manipulation and access should be considered in priority. Existing UCC writing tools require many hours and efforts in learning functions to use, so easy access was hard to be attained. Therefore, even beginner should be able to use it easily by making the interface for manipulation easy.

Second, consideration should be given to produce interactive contents that users can use in various fields. Mobile based SNS-UCC writing tool should be developed to be utilized in various fields like UCG(User Created Game), relay novel, music video and advertisement etc as well as existing UCC. The more fields the tool can be utilized in, the more users will use it, and the more diverse types of profit models will be created.

Third, mobile based SNS-UCC supports easy transfer using MMS(Multimedia Message Service) and mobile web. MMS and mobile web should be supported simultaneously by considering the internet environment of touch phone and smart phone that are drawing attention recently as well as existing cellular phone and finally, convenience should be improved through interlinking with PC internet.

Fourth, mobile based SNS-UCC should be designed so that it can be used in any mobile device through compatibility with various platforms. Mobile based SNS-UCC should be able to use writing tool in various mobile devices like smart phone etc represented by iPhone and Android phone as well as the cellular phone that people carry the most.

3 Cases of Mobile Based SNS-UCC Writing Tool Development

Mobile based SNS-UCC writing tool is under development through support by 2009 IT Excellent Technology Supporting Project by applying the principle of mobile based SNS-UCC writing tool designing. "SNUC" the mobile based SNS-UCC writing tool is under development by reflecting the 4 designing principles that we went over above, and has the following characteristics:

1) Contents can be produced “anytime anywhere”.
In existing UCC production, space that accommodates use of PC was required, thus facing limitation of time and space. However, "SNUC" is of mobile base of easy portability, thus offering the advantage of being able to be produced with no limitation of time or space.
2) Transfer of "emotion" through story telling is possible.
Existing UCC was used often in editing moving picture or photo. "SNUC" can make theme story by utilizing the photo, moving picture or music etc of its own in mobile device, so it can create more emotional story. Also, "SNUC" should be made so that it can produce contents with various story with the conception that people remember contents with story longer. "SNUC" supports production of contents with story by offering structure of story with themes by the situation called theme template.

3) SNS that can be shared with “anyone”.
In the case of existing SNS, information of its own should be entered and saved individually by accessing the site through PC internet. But, mobile device already contains all the information of my own, no entering of information is required. In addition, mobile device is also a tool most widely used in communicating with and establishing relationship with people.
4) Can produce contents “easily and simply”.
To produce UCC in the past, developers should learn and use complicated tool, thus requiring long hours in learning the tool and facing hardship in access. However, "SNUC" is considerate enough to accommodate completion of contents only through providing of template in a completed form with themes by the situations and simple modification.
5) Offers the extensibility “that can be used in various fields”.
"SNUC" is designed to be used in various fields like novel, film, music video, advertisement, marketing and drama etc as well as the function of SNS-UCC. Owing to the function of storytelling, "SNUC" can be applied in various types like UCG(User Created Game), UMC(User Modified Contents" and URC(User Recreated Contents) etc in addition to UCC.

4 Conclusion

Yet, the researches and developments of mobile based SNS and UCC are at beginning stage compared with PC based internet. However, many experts and companies are conducting research and development with significant interest in SNS and UCC.

It is anticipated that once mobile based service with environment suitable for the supply of SNS is presented, not only SNS field but also UCC and Web 2.0 service on such basis will also evolve one step further. In such sense, it is important to establish the principles for the development and designing of mobile based SNS-UCC writing tool.

Development of "SNUC" will play the role of uplifting the reachability of users as writing tool that can be used for easy and simple production in mobile based SNS market which is still at beginning stage.

References

1. Ahn, S.-H., Song, S.-M.: Status and Development Prospect of UCC Service. Korea Contents Society General Academic Convention Journal 5, 691–697 (2007)
2. Kim, S.-Y.: Trend and Prospect of Mobile SNS Service. Korea Communication Society Journal 26, 19–25 (2009)

Project Sonology: An Experimental Project Exploring the Possibilities of Sound and Audio as the Primary Element of Interactive Entertainment

Shih-Han Chan, Dae Hong Kim, Eugene Massey, Katelyn Mueller, and Fadzuli Said

Entertainment Technology Center, Carnegie Mellon University,
700 Technology Drive, Pittsburgh, PA 15219, USA
{Shihhan.Chan,Daehong75,Eugene.H.Massey}@gmail.com
{Katelyn.Lisa.Mueller,Fadzuli.Said}@gmail.com

Abstract. The goal of the project is to show that audio can successfully be the primary element of interactive entertainment by delivering pure audio experiences that demonstrate both the creative potential and emotional power of an audio experience. We develop two proofs of concept with the technical foundation supported by prototypes. The core technology is a combination of a 3D game engine and an audio engine used to build sound environments. The interactions are based on 3D trackers and surround sound headphones.

Keywords: Interactive Audio, Virtual Environment, 3D Sound.

1 Introduction

Project Sonology is a student pitched project of the Entertainment Technology Center at Carnegie Mellon University in 2008 whose goal is to show that audio can play the major factor in interactive entertainment by creating pure audio experiences that are unique and present both the creative potential and emotional power of an audio experience. We create a flexible technical foundation supported by multiple prototypes demonstrating the technology that would help shape the experiences we would eventually deliver. With the technical foundation in place, we work on two proofs of concept as the final deliverables: The Story World and The Music World.

We combine a 3D game engine and an audio engine as the core technology to create 3D sound environments that can be navigated and interacted with intuitively. The interactions are based on 3D trackers and surround sound headphones. Guests perform actions and explore different parts of the environment.

2 Platforms

FMOD audio engine and OGRE 3D game engine are chosen as the fundamental software. FMOD gives sound designers the option to create complex audio event behaviors. OGRE comes with 3D math and game utilities that help to speed up development.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 431–433, 2010.

FMOD and OGRE allow us to divide up and modularize the programming work so that different team members can work on different aspects simultaneously. On the other hand, Polhemus 3D trackers and surround headphones are selected as core hardware platforms in the final proofs of concept. The surround headphones we use have speakers in the front and back for an added dimension.

3 Prototypes

In the early stage of the project, we developed a series of prototypes. From that we received valuable feedbacks directing us to revise and polish the work. The various prototypes that help define and demonstrate the technology foundation embrace:

- Head tracking to allow the guest to walk around and rotate his heads while listening to different aspects of a virtual 3D environment.
- Hand tracking to allow the guest to reach out sounds and then move them to new positions in the virtual 3D space.
- A fly catching game to integrate head and hand tracking into a system.
- Webcams with OpenCV to test out possible techniques for body tracking.
- DSP Plug-ins to simulate binaural effects.
- Speech recognition to implement auditory menus.
- Directional focused sound to assist the guest in locating a sound source.

4 Proofs of Concept

4.1 The Story World: Jack and the Beanstalk

Since audio is an ancient platform for story-telling, we want to deliver a story by using our interactive audio-only concept. The goal is to find a good balance between immersion and interactivity. An interactive version of “Jack and the Beanstalk” is made where the guest assumes the role of Jack. By using Jack’s mom as the guide, the guest would feel that he is interacting with another character.

A variety of simple interactions with sound are applied in many places such as grabbing the beanstalk, catching the hen and chopping the beanstalk. We also utilize the voice recognition technique in which the guest is requested to sing and match the tone that the golden harp plays. Yet some guests have had a hard time finding objects, we add a “Hot & Cold” feature, by which our program would guide the guest to a certain object by providing information of the distance from the target to the guest.

4.2 The Music World: Whip It

Our other proof of concept is an experience based around listening to music. The guest would hear a song, in this case Devo’s Whip It, and then the song would break into pieces at which point the guest would have to reconstruct the song to complete the experience.

The main interactions with this experience are finding sounds and then playing with them. We add a sonar feature that makes a beeping noise to increase in intensity as the guest gets closer to a sound that he needs to collect. Once the guest finds a sound, he can play with it. For instance, when the guest finds the guitar sound, guitar noises would be produced if he plays “air guitar.” Most guests do make these kinds of motions without any prompting.

This experience is meant to be a tutorial for our technology and it works very well in this regard. It is also designed to be simple and fun to boost the guest’s confidence.

5 Observation and Lesson Learned

People are happy with the simplest aspects of our technology. With a complex and new experience, very simple interactions that are intuitive and recognizable are more fun than more complicated combinations of the interactions.

Another lesson we learn is that it is very difficult to place a sound in 3D with enough information that a person can tell naturally where the sound is. When the sound is behind the guest, we apply a filter that blocks out all the high frequencies; when it is on the side, we pan the sound, attenuate volume and apply a Doppler effect for the distance, whereas in actual reality, sound is bouncing off walls and being absorbed by the material of the carpet. We also learn that the human hearing system is extremely complex. Our brain is automatically filtering out all the sounds that are not vital to allow you to listen to what is being said. We have attempted to solve this problem within the scope of the technology we have by adding story and context to help fill up gaps in information that our sound cannot convey. We moreover give other kinds of feedbacks like sonar beep noises to show the proximity of hands to objects and move sounds around in a certain pattern so that the guest can easily position the sounds. Though how to make the sound localization more clear is another course for us, our attempts meet with enough success that a guest can pick up quickly how to locate sounds in our world and proceed on to the rest of the experience.

6 Conclusion

As previously stated, our goal is to prove that audio can be the primary element of an interactive experience. We believe that our project is successful in demonstrating this, but that our proofs of concept have only scratched the surface of what is possible in this realm. For example, we only make “Jack and the Beanstalk” interactive, but this proves that the same can be done for any story; we only break apart “Whip It” and put it back together, but the same can be possible for any song in any genre. In addition, this technology can be adapted for use with mobile devices like cell phones, MP3 players, as an interactive museum installation, a musical instrument or even as an educational tool.

We hope that the lessons learned in our project and what we have achieved would motivate future audio-only projects to reveal more potentials of audio and sound with interactivities.

Multipresence-Enabled Mobile Spatial Audio Interfaces

Owen Noel Newton Fernando[1], Michael Cohen[2], and Adrian David Cheok[1]

[1] Keio-NUS CUTE Center, National University of Singapore, Singapore 119613
{newtonfernando,adriancheok}@mixedrealitylab.org
[2] Spatial Media Group, University of Aizu, Aizu-Wakamatsu, Japan 965-8580
mcohen@u-aizu.ac.jp

Abstract. Mobile telephony offers an interesting platform for building multipresence-enabled applications that utilize the phone as a social or commercial assistant. The main objective of this research is to develop multipresence-enabled audio windowing systems for visualization, attention, and privacy awareness of narrowcasting (selection) functions in collaborative virtual environments (CVEs) for mobile devices such as 3rd- and 4th-generation mobile phones. Mobile audio windowing system enhances auditory information on mobile phones and encourages modernization of office- and mobile-based conferencing.

1 Introduction

Vision and audition are the two main human senses for obtaining information about the outside world, and rich interfaces need both modes (at least!). Visual windowing systems allow multiple and multiwindow applications to share display resources; audio windowing systems, in analogy to graphical windows, can bring order to a cacophony of multiple simultaneous sound sources. Audio windowing can be thought of as a frontend, or articulated user interface, to a system with a spatial sound backend. Audio windowing systems are especially appropriate on small screen devices where there limited space for information and when control changes are not needed.

Researchers are working on interactive spatial audio to support synchronous groupware like teleconferences, either desktop applications or mobile device applications. For instance, Herder and Yamazaki [2] describe a networked virtual environment with 3D audio to support a chatspace while Aoki et al. [3] present a mobile audio space intended for use by gelled social groups. To the best of our knowledge, there is no existing presence awareness systems that can handle multiply present sources and sinks. The narrowcasting attributes [1] presented here suggest an elegant solution for such multipresence environments.

The suite of inclusion and exclusion narrowcasting commands for sources and sinks are like analogs of burning and dodging (shading) in photographic processing. The duality between source and sink operations is tight, and the semantics are identical: an object is inclusively enabled by default unless, a) it explicitly excluded (with `mute` or `deafen`), or, b) peers (either `self`-designated or non

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 434–436, 2010.

self-designation of the selected object) are explicitly included (with select [solo] or attend) when the respective avatar is not. We have developed a mobile phone interface which allows users to multicast their voice to multiple receivers and control crowded soundscapes using these narrowcasting operations. Programmed with J2ME, our application runs on NTT DoCoMo iappli mobile phones.

2 "ı·Con": Mobile Device Multispace Interface

Featuring selectable icons with one rotational and two translational degrees of freedom, the "ı·Con" 2.5D dynamic map interface is used to control position, sensitivity, and audibility of avatars in a groupware session. Its isosceles triangle icons are representations of symbolic heads in an orthographic projection, including narrowcasting attributes.

We have developed symbolic representations of narrowcasting operations (as shown in Figure 1) for the mobile interface. In our mobile application, narrowcasting attributes graphical displays are triply encoded— by position (before the "mouth" for mute and select, straddling the "ears" for deafen and attend), symbol ('+' for assert '-' for inhibit), and color (green for assert red, yellow, and orange for inhibit). The attributes are not mutually exclusive, and the encoding dimensions are orthogonal. For instance, a sink might be first attended, perhaps as a member of some non-singleton subset of a spaces sinks, and then later deafened, so that both attributes are simultaneously applied.

The select and attend attributes are denoted by characteristic features. If any avatar has been selected, non-selected avatars are implicitly muted if their self-designation state is the same as the selected avatar. For example, if a self-designated avatar is selected, only other self-designated avatars are implicitly muted. In the same manner, if any attended avatars exist in a given space, non-attended avatars are implicitly deafened if they are in the same class (self or non-self) as an attended avatar. Orange color is used to represent these implicit effects, implicit mute represented by a '-' before the mouth and implicit deafen by '-'s straddling the ears.

A unique feature of our system is the ability of a human pilot to delegate multiple avatars simultaneously, increasing quantity of presence; such multipresence enables one to overcome some fundamental constraints of human condition. Our interface will allow each human user to have sources and sinks present in multiple conversations at once. Users will be able to receive multiple audio streams and transmit their voices to multiple recipients, controlled according to custom preferences. Multiple sources are useful, for instance, in directing one's remarks to specific groups, decreasing the granularity of audibility control. Multiple sinks are useful in situations in which a common environment implies social inhibitions to rearranging shared sources like musical voices or conferees, as well as individual sessions in which spatial arrangement of sources, like the configuration of a concert orchestra, has mnemonic value.

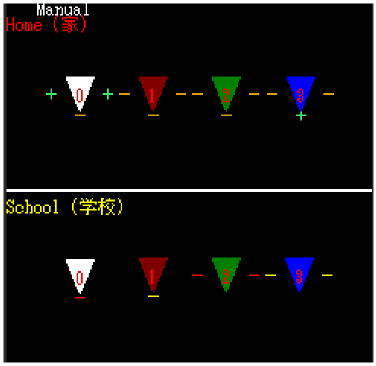


Fig. 1. Narrowcasting attributes graphical displays: In the Home space, avatar #0 is `attend`ed, so its complement (comprising all the other avatars) is `deafen`ed; and avatar #3 is `select`ed, so its complement is implicitly `mute`d. In the School space, avatars #0 & #2 are respectively `mute`d and `deafen`ed by self while avatars #1 & #3 are respectively `mute`d and `deafen`ed by others.

3 Conclusion

The basic goal of this research is to develop rich idioms for selective attention, presence, privacy, and narrowcasting for presence awareness applications for a nomadic device like a mobile phone. A multipresence feature using these idioms encourages users to install avatar representatives of themselves in several places and spaces at once. Spatial sound enhances the conferencing capabilities in collaborative virtual environments. Multipresence-enabled systems will enhance office communication and make it easier to coordinate work groups. The system also enhance human communication, allowing users to interact with friends, family, and colleagues "anytime anywhere." We expect that commercial development of this research will involve partnerships with network providers, who might license such technology to offer to their subscribers as an added value service.

References

1. Cohen, M.: Exclude and include for audio sources and sinks: Analogs of mute and solo are deafen and attend. Presence: Teleoperators and Virtual Environments 9(1), 84–96 (2000) ISSN 1054-7460
2. Yamazaki, Y., Herder, J.: Exploring spatial audio conferencing functionality in multiuser virtual environments. In: 3rd International Conference on Collaborative Virtual Environments, pp. 207–208. IEEE Press, San Francisco (2001) ISBN 1-58113-303-0
3. Aoki, P.M., Romaine, M.H., Szymanski, J.D., Wilson, T.D., Woodruff, A.: The mad hatters cocktail party: a social mobile audio space supporting multiple simultaneous conversations. In: Proc. of the Conference on Human Factors in Computing Systems, Ft. Lauderdale, Florida, USA, pp. 425–432. IEEE Press/ACM Press (2003)

Fluxion: An Innovative Fluid Dynamics Game on Multi-touch Handheld Device

Chun-Ta Chen, Jy-Huey Lin, Wen-Chun Lin, Fei Wang, and Bing-Huan Wu

Entertainment Technology Center, Carnegie Mellon University, Pittsburgh, USA
fluxion.cmu@gmail.com

Abstract. We explore the possibility of implementing real-time fluid simulation on iPhone to create an innovative game experience. Using fluid dynamics and water tri-states as game mechanics, players can manipulate fluid and solve puzzles through the unique input controls of iPhone, such as accelerometer and multi-touch. We implement particle-based fluid simulation and integrate our particle system with a physics engine, Box2D, to realize the interactions between particles and rigid body. The playtest showed that Fluxion is not only a fun game, but also educational since it provides players the basic concepts of how fluid behaves in the real world.

1 Introduction

Water plays an important role in our daily lives, and it is always fun to play with it. Nowadays, fluid simulation has become an increasingly popular topic in computer graphics for generating realistic animations of water, smoke and related phenomena. With the novelty of iPhone, we aim to create an iPhone game using fluid dynamics as game mechanics, so that players can use iPhone featured input controls to manipulate fluid and solve puzzles.

(a) (b) (c)

Fig. 1. (a) The tri-states of water: gas, liquid, and solid. (b) The according liquid behavior when the accelerometer senses tilting right motion. (c) The item menu, showing all items that players can use in the current level.

We develop an immersive game experience and explore the possibilities of innovative game mechanics based on fluid dynamics. There are three main features in our game. First, our game mechanic is based on fluid dynamics. The biggest difference from others is that we integrate the water tri-states into our game (Figure1.(a)). Players

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 437 – 439, 2010.

can use tools such as heater, cooler, freezer etc. to play with state changes between gas, liquid and solid. Second, iPhone's input controls have a great potential for intuitive and interesting game interaction. The 3-axis accelerometer senses the orientation of the phone and changes the content in the game accordingly (Figure1.(b)). For example, with the accelerometer, players can tilt the iPhone to move or to speed up the fluid flow. Third, our game involves a lot of conceptual challenges for players to utilize fluid features to solve puzzles. We design tools to help players solve problems (Figure1.(c)). For example, players can place a heater to turn water into gas or place a freezer to turn it into ice.

2 Game Design

In fluxion, the objective of each level is to collect enough fluid particles to reach the goal. Players can tilt the iPhone to control the gravity of fluid. They can also use different items to perform state changes between gas, liquid and solid, or change the flowing direction to pass through obstacles and solve puzzles. There are some other interesting features such as using fingers to block the fluid movement and using multi-touch to control camera movements.

Fluxion is a game introducing tri-state features. It includes solid, liquid, and gas. As shown in Figure 2, three elements form a bi-directional chain where each element can be transformed into its neighbor states.

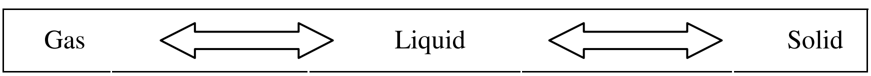
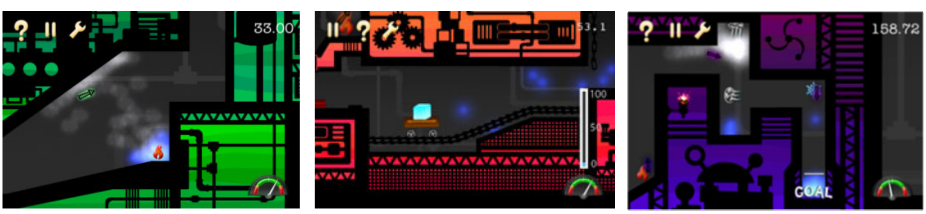


Fig. 2. State Transitions

Fluxion provides seven kinds of items that help players pass each game level. These items are freezers, coolers, heaters, and four kinds of fans blowing (upward, downward, rightward and leftward)(Figure3.).



(a) (b) (c)

Fig. 3. (a) A heater turns water into gas. (b) Water is turned into an ice cube so that the cart can carry it. (c)Fans blow gas away.

3 Technical Approach

Fluid dynamics is a wide topic. There are a lot of approaches to simulate realistic fluid in computer graphics. In general, these approaches can be divided into two groups: grid-based and particle based fluid simulation. In our system, we apply Smooth Particle Hydrodynamics (SPH) [1], which is a particle-based simulation method.

An important issue is how to implement fluid transformation between different states. We set a temperature variable to control the state change. When the temperature of a particle reaches a certain threshold, the particle will change its state. For freezing water into ice, we use a physics engine (Box2D) instead of our particle system to form a rigid body as an ice and handle the interactions between particles and the rigid body.

We optimize the OpenGLES drawing pipeline and our rendering algorithm by using texture-based rendering instead of grid-based rendering, which enormously improves the performance. Fluxion has been performed on iPhone 3GS at 30 frames per second with 150 particles.

4 Conclusion

Beyond recreational purpose, Fluxion is also educational. Players can learn the fluid physics through playing Fluxion. It completely shows the property of fluid dynamics and the tri-state of water: solid, liquid and gas. In addition, Fluxion demonstrates phenomena of fluid particles: water splashes while colliding to walls, gas always flows up, and ice cubes always floats on top of water. Moreover, our game allows players to perform logical thinking in the process of solving the puzzles using the knowledge they learned from the game. All in all, Fluxion gives players a basic idea of fluid behaviors in the real world.

Another achievement of our work is that we implement a real time fluid simulation on iPhone. Since iPhone has limited computation power, we manage to utilize the most efficient way to simulate fluid. Reducing unnecessary function calls and optimizing the algorithm make our game run at an acceptable frame rate. Our game takes advantages of iPhone features that make it more intuitive for players to observe and manipulate water particles around on a portable device rather than pressing arrow keys and watching the simulation on a still monitor.

Reference

1. Müller, M., Charypar, D., Gross, M.: Particle-Based Fluid Simulation for Interactive Applications. In: SIGGRAPH Symposium on Computer Animation (2003)

Robotic Event Extension Experience

David Obdrzalek

Charles University in Prague, Faculty of Mathematics and Physics
Malostranske namesti 25, 118 00 Praha 1, Czech Republic
david.obdrzalek@mff.cuni.cz

Abstract. In this paper we present our experiences from extending the Eurobot contest for students of the age up to 30 by a category for pupils up to 18. We show two different models of the extension and present our experiences acquired after implementing them in 2008 and 2009.

Keywords: Autonomous robot contest, Education, Edutainment, Entertainment robotics.

1 Introduction

Robotic contests are nowadays popular. At the same time, the increasing power, abilities, and "user-friendliness" as well as decreasing price of components usable for building robots makes them available to broad public, and more and more for younger and younger robot builders. In this paper, we present the Eurobot Autonomous Robot contest and how we have extended it in the Czech Republic also for younger participants. For more details about the rules, participation conditions, discussions and reasonings see [1] and [2].

The following text is organized as follows: in Section 2, we present Eurobot contest. In Section 3, a contest for younger participants is introduced as Eurobot Junior and in Section 4, we present another new category called Starter. In Sections 5 and 6 we give our experiences from the two years when Eurobot + Eurobot Junior (2008) and Eurobot + Starter (2009) contests have been organized in the Czech Republic. Section 7 concludes the paper.

2 Eurobot Contest

Eurobot is an international autonomous robot contest for young non-professional roboticians of the age up to 30 years. Their task is to build an autonomous robot which is able to perform actions defined by the rules. Then, the builders meet for matches with other teams and their robots. In the matches, two robots from two teams compete on a playing field. The core technical rules of the Eurobot contest are: autonomous robots (no remote control allowed); indoor robots with limited size (roughly a 30 cm cube); game on a table (roughly 2 x 3 m); little time for one match (90 sec); fair-play spirit of the game; team work.

While these basic technical rules remain the same, the rest is changed every year; it is therefore easier for new participants to start with Eurobot in contrast

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 440–442, 2010.

with other contests where the rules are fixed and it takes years to reach the top level (see e.g. RoboCup initiative [5]). The changing part of the rules defines the main topic, playing elements etc.: during the years, the robots had for example to build structures from wooden cubes, collect coloured balls, sort waste, or lay down skittles (see the Archives section at [1]).

3 Trophées de Robotique / Eurobot Junior Contest

The Eurobot contest started as a single-category contest. In 2006, Eurobot organizers asked organizers of another contest, Trophées de Robotique, to join Eurobot and open it also for other countries under the name of Eurobot Junior (see [3] and [4]). This category is limited for participants of the age up to 18 years or the end of their secondary school studies. The task is simpler than in the main Eurobot contest; the most significant difference is that the robots are not autonomous but remote controlled. But, the rules are by intention prepared independently, and Junior rules share with Eurobot only the core (see Sec. 2).

4 Starter Contest

In 2007, Czech organizers of Eurobot National Cup organized only Eurobot contest. In 2008, they prepared Eurobot Junior too. For 2009, they decided to run the contest for the younger category like in 2008, however to ease the organization, they decided to use another model: to use Eurobot rules and not to make two different contests. The only change done to Eurobot rules was the non-autonomy of Starter robots (and attributes related to this), the rest of the rules stayed intact. Therefore, the two categories look very similar, because the playing field is the same and most other visual attributes too, and the match goals for the robots are exactly the same. The Starter rules specification can be found at Eurobot Czech web pages (see [2]).

5 Eurobot + Eurobot Junior in the Czech Republic, 2008

For the visitors, both contests were interesting and the whole event was “diverse and colourful” (because of two different playing fields), however it was quite difficult for them to understand the two different rule sets.

The Eurobot and Junior teams stayed separate and concerned only on their respective category matches.

For the organizers, adding Eurobot Junior meant organizing two completely different contests even some external attributes were seemingly similar. For example, the implementation required to translate two sets of rules from English to Czech, build two different playing fields, train two groups of referees etc. During the debriefing, the organizers had to confess they underestimated how different the two contests were to organize.

6 Eurobot + Starter in the Czech Republic, 2009

The nearly same rules for Eurobot and Starter in 2009 created a "compact" event, which was easier for the public to follow and understand. The public also appreciated the kids took part in the same technical and scientific activity as the older participants.

The Eurobot and Starter teams merged together and examined in detail the results of work of the other teams regardless on the category, because all participants had the same goals and therefore the two groups had the same knowledge of what to expect and what to look for. The organizers consider it as a great result, fulfilling one of the goals of Eurobot Association – "to foster scientific and educational exchange between young amateur robotic fans".

For the organizers, making Starter did not bring any additional work to Eurobot, in general the only difference to "Eurobot only" was just bigger number of teams (Starter + Eurobot).

7 Conclusion

In this paper, we have shown the evolution of the Eurobot contest between 2007 and 2009. The experience clearly shows several conclusions:

Firstly, the idea to extend the contest also to involve younger participants was good and eligible.

Secondly, adding the younger category did not only allow younger people to take part and learn from the experienced contestants in the main category, but also has shown that older participants can (and do) gain from the youngsters. Originally, the category for younger participants was introduced because the younger scholars were afraid of competing with university students. It has shown that in many cases this fear is unjustified and even more, in some cases the older students from universities have learnt from the younger scholars!

At third, the organizers of Eurobot contest in the Czech Republic have found out that for small- and middle-sized national cups (currently all except France), the contest model used in 2009 serves better than the model used in 2008 when the two age categories were separated. It might not be the case for big events where plenty of teams participate, because it is not possible for the organizers to run it at the same time on the same place anyway. But for all other national organizing groups where the two categories could be organized together, the advantages are clear and expressive.

References

1. Eurobot Association: Eurobot Autonomous robot contest, http://www.eurobot.org
2. Eurobot Association: Eurobot Czech Cup, http://www.eurobot.cz
3. Eurobot Association: Eurobot Junior, http://www.eurobot.org/eng/junior.php
4. Planète Sciences: Trophées de Robotique, http://www.planete-sciences.org/robot
5. The RoboCup Federation: RoboCup international research and education initiative, http://www.robocup.org

A Sound Engine for Virtual Cities

Shih-Han Chan, Cécile Le Prado, Stéphane Natkin, and Guillaume Tiger

CEDRIC, CNAM, 292 Rue Saint-Martin
75003 Paris, France
shihhan.chan@gmail.com, cecile.leprado@free.fr,
stephane.natkin@cnam.fr, tiger.guillaume@gmail.com

Abstract. This paper is a position paper to specify and implement a general-purpose sound engine for virtual cities. The work is motivated by the project Terra Dynamica funded by the French government. We present a state of the art of the virtual urban sound spaces emphasizing various types of virtual cities and their relationships to auditory space. We then discuss the choice of a sound engine, sound spatialization and scene description languages as ongoing works.

Keywords: Virtual City, Virtual World, Sound Engine, Soundscape.

1 Introduction

The purpose of this research is to study the dynamic of sound in virtual cities and to propose a sound engine solution that allows sound designers to organize virtual urban soundscapes in an interactive 3D environment. This work integrates the continuous research of the CEDRIC laboratory on sound and virtual environments.

Our work is motivated by the project Terra Dynamica, a real-time simulation of the city with avatars and non-player characters (individual characters, crowds, vehicles, flows). This project consists in both an adaptable architecture and a granular approach from individual to group. In the deliverable, real-time sound scenes are built in 3D through an open soundscape editor including procedural behavior. While sound plays different roles within different scenarios, this platform in development allows sounds to provide information, characters' locations for instance, in various fields of applications such as safety exercises and art design.

2 State of the Art

In the field of interactive media, considering urban representations, the audio content is closely related to the spatial organization of the city. In addition, nowadays' digital representations are often derived from soundscape studies carried on during the sixties.

Urban sound representations come from acoustic ecology and noise pollution matters. This mostly takes the form of soundmaps which consist of empirical auditory content overlaid on mapped space and, sometimes, tend to cross-media applications. Furthermore, considering the city as a designed space, its relationship to sound can be

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 443–445, 2010.

understood as a dynamic interaction, first through the analytic and creative paradigms of the soundscape theory, then in regard of the architectural digital acoustic tools and finally as a real-time construction within video games.

2.1 Dynamics of Urban Auditory Environments: Soundscapes, Architectures and Video Games

Audio, in soundscape theory, architecture and video games, is characterized by its strong relationship to spatial organization. The soundscape theory describes the auditory environment as a perspective (background, foreground and soundmarks) thus creating multidimensional dynamic objects. This perspective is echoed in architectural auditory virtual representations which provide advanced acoustic simulations calculating propagation fields (direct, diffuse, and critical). Finally, some video games offer large, real-time, interactive, 3D urban environments in which sound design strategies attempt to create a credible environment rather than a simulation.

2.2 Usage of Sound in Virtual Cities

We select typical applications of sound in virtual cities. Google Earth is the virtual earth representation; Second Life is also a multipurpose virtual world; Prototype is an action-adventure video game (Radical Entertainment / Activision) which uses an original dynamic sound map of New York City; Locustream is an "openmike" soundmap used for many interactive purposes; Soundwalk allows combining a localized sound recording with the real sound of the city. Figure 1 shows the use of sound in each application.

SOUND FUNCTIONS \ MEDIA	GOOGLE EARTH	SECOND LIFE	PROTOTYPE	SOUND WALK.COM	LOCUS STREAM
IMMERSION	**Static recordings**	**Public spaces, sound art works and private spaces**	**Music, surround sound, realtime spatialization**	**Ambient sound, read text**	**Realtime audio streams**
NAVIGATION HELP			**3D sound**	**Audio guide**	
INFORMATION	**Relative to listenable audio files**	**Positioning objects (3D sound), relative to audio content**	**Character behaviors, density / proximity of agents**	**Historical and culture information, localization**	**Relative to audio stream broadcasts**
SOCIALIZATION		**Concerts, adding sounds to avatars, private audio broadcasts**			**User community**
STORYTELLING			**Interactive music, dialogues**	**Sound path divided into sequences**	
REUSE OF DATA					**Utilization of streams as sources**

Fig. 1. Intersection of different types of virtual urban media and sound functions (red text shows functions for which users may bring content)

In all applications, the construction in layers reminds of the classical theory of soundscapes: background, foreground, and soundmarks. The dynamic organization of the soundscape is according to the needs created by the gameplay or rules of the media.

3 Conclusion

Most of sound game engines have a three level architecture. The lower level includes sound synthesizers and real-time filters that can compose and be applied to a sound stream which asks for a balance between procedural sound design and wavetable playback. We choose Collada as the formal representation of the virtual space on which the intermediate level of the audio architecture relies. The definition of a sound extension of Collada is being made based on the scene language of MPEG4. The design of an efficient high-level editor is the core of our project, and is based on the analyses of several experiments and usages in the Terra Dynamica project.

We conduct various urban recordings to implement an interactive soundmap as the preliminary work. On top of that, the spectral analysis of the recordings provides data about the layers composing the urban soundscape. In our first experiment, Fmod is the lower level tool integrated with the 3D rendering engine – Unity3D. We will next explore other possible combinations of various sound engines (e.g Wwise, Playall) and rendering engines (e.g Ogre) with the integration of Collada sound scene.

Considering the large number of audio sources in urban virtual environments, spatialization problematic will be included in our future works as it may bring solutions regarding level of detail, source clustering and reverberation.

While the project Terra Dynamica requests audio uses within five main fields of applications (game and art, urban design and architecture, safety analysis, city transportation, proximity services), it will also provide us some visual data and basic requirements (e.g. path finding) from which we will be able to define more precisely our sound engine architecture.

References

1. Collins, K.: Game Sound. MIT Press, Cambridge (2008)
2. Wrightson, K.: An introduction to acoustic ecology. Soundscape the Journal of acoustic ecology 1(1) (2000)
3. Funkhouser, T., Tsingos, N.: Sounds Good to Me! In: Computational Sound for Graphics, Virtual Reality, and Interactive Systems. In: SIGGRAPH 2002 Course Notes (2002)
4. Veneri, O.: Architecture d'un intergiciel pour la creation sonore dans les jeux video. Phd thesis, CNAM (2009)
5. Scheirer, E., Lee, Y., Yang, J.: Synthetic Audio and SNHC Audio in MPEG-4. In: Puri, A., Chen, T. (eds.) Advances in Multimedia: Signals, Standards, and Networks. Marcel Dekker, New York (2000)

NetPot: Easy Meal Enjoyment for Distant Diners

Zoltan Foley-Fisher, Vincent Tsao, Johnty Wang, and Sid Fels

Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada
{zoltan,vtsao,johnty,ssfels}@ece.ubc.ca

Abstract. We capture key factors of a group meal with communication and interface technologies to make a meal more enjoyable for diners who cannot be collocated. We determined three factors of a popular group meal, Chinese hotpot, that are essential for a group meal experience: interacting as a group with food, a central shared hotpot, and a feeling that others are nearby. We developed a prototype system to maintain these factors for an online meal with remote friends. Our technique is of interest to designers creating technology for isolated diners.

Keywords: Augmented Reality, Hotpot, Online Meals, Social Networking.

1 Introduction

Family or friends often enjoy meals together, but sometimes not everyone can join the meal. This is particularly true as older people live independently, younger people live alone [1] and working people travel more and work from afar. In these cases, one or all the group has to eat alone – an experience that is not as enjoyable as a group meal.

Ubiquitous, high bandwidth and inexpensive communications technology offer a solution. Telepresence technologies [2], like Skype, may have interaction limited to audio and grainy video, but we can build on these technologies to create a system that captures the factors of an enjoyable group meal so that remote diners can join the fun.

To our knowledge, we are the first researchers to use targeted interactions to focus only on communicating important factors of group meals with simple augmented-reality technologies rather than trying to recreate a highly realistic tele-eating experience. Maintaining the factors with our prototype ultimately improved diner's enjoyment when sharing a hotpot meal with remote friends. Our system can be made cheaply and is easy to set up. Until truly realistic telepresence technologies are available, our approach may well be the most enjoyable way to eat with remote friends.

2 Missing Meal Factors

In our study, we focussed on Chinese hotpot. Hotpot is a communal meal eaten in East Asia, and is popular among Canadian-Asian communities in Vancouver. In hotpot, a group of friends or family members simmers a central pot of stock at the dining table

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 446–448, 2010.

where each person uses their chopsticks to select items to cook from an assortment of plates of food around the table such as sliced meat, vegetables and seafood. Hotpot is a social meal that takes a long time to eat and is often enjoyed on special occasions. We surveyed diners and observed three factors to be important in this meal:

1. **Interacting as a group with food.** Some diners adopt roles, such as the person in charge of putting in specific foods. There is often a formal order when food is taken out of the pot. Some diners test the food to tell others if it is cooked.
2. **Central shared hotpot.** Hotpot diners all share the same pot. Diners rarely turn away from the hotpot. The meal lasts a long time. Even when talking to others, diners focus their attention on the food in the pot.
3. **Company, a feeling that others are nearby.** Even though diners do not pay a lot of attention to others, they are nevertheless aware of them in their periphery.

By maintaining these factors, we expect that meals with remote diners can be made more enjoyable. We also aim to select the least complex and expensive technologies that will result in significant improvements to meal enjoyment. Teleconferencing software, such as Skype, cannot recover all the missing factors alone. We developed low-fidelity prototypes to determine that designs to recover meal factors should: Adopt avatars to represent remote diners, Acknowledge that diners are not concerned with what remote diners are eating, Involve simple interactions, Use readily available and inexpensive components, and Make sure audio is included.

3 Prototype

For our prototype, half of a traditional Yuān Yāng Guō, or "Mandarin Duck”, hotpot is used for cooking, while the other half serves as a display and tracking surface to interact with remote diners. The interactive surface of the pot is a white space on which images are projected, and the movements of the diner’s chopsticks within this space are tracked with a video camera. An induction stove is used as a safe and compact heating element. This is shown in Fig 1.

Up to three virtual food items, such as gailan (a type of broccoli), fish balls and tofu, appear on the interaction surface as if dropped into the pot. The food floats around as it slowly gets “cooked”. The food initially appears raw (meats are red/pink in color), and changes to a cooked appearance as it stays in the pot. If food has been in the pot too long, it appears “overcooked”. A diner can move a piece of food by placing the tip of the chopstick on it, and then moving the chopsticks. For simplicity, the food cannot be dropped once picked up, but only placed into a “container”.

Three diners are represented by avatars at the edge of the interactive surface. A diner is fed when food moves close to their avatar. Each diner has a “satisfaction bar” that is incremented each time they are fed. Cooked food increments the bar, overcooked food decrements the bar. Diners are notified of the names of remote diners who feed their avatar. The avatar enters a sleep state if remote diners don’t interact with their hotpot.

The system augments a typical hotpot arrangement with a computer, overhead projector and webcam to track the chopsticks and runs over the internet using our software and Skype™.

Fig. 1. Inductive hot-plate and the NetPot system of projected remote diner actions, chopstick tracking cameras and Skype audio teleconferencing

4 Conclusion

We determined three meal factors missing when diners eat Chinese hotpot alone. We assumed that if these factors were maintained for a tele-eating experience, then diner's meals would be more enjoyable without trying to recreate the exact meal experience using high-end teleconferencing. We targeted interaction designs with readily available augmented reality technology to keep the missing factors in an online meal with remote friends.

We evaluated a prototype of the design – NetPot – to assess how much more enjoyable the meal is than when sharing the meal with a low quality telepresence system. We maintained the meal factors, although not completely. The meal was more enjoyable, but it is not clear if it was because the factors were there or because the interactions we chose were enjoyable.

Until telepresence systems provide complete realism and are affordable, a careful treatment of the important meal factors for remote diners with targeted interactions may lead to more enjoyable shared dining even when friends and family cannot be together.

References

1. Sellaeg, K., Chapman, G.E.: Masculinity and food ideals of men who live alone. Appetite 51(1), 120–128 (2008), doi:10.1016/j.appet.2008.01.003
2. Sheridan, T.B.: Musings on telepresence and virtual presence. Presence: Teleoperators and Virtual Environments 1(1) (1992), `http://portal.acm.org/citation.cfm?id=128956`

Glasses-Free 3D Display System Using Grating Film for Viewing Angle Control

Masahide Kuwata and Kunio Sakamoto

Department of Intelligence and Informatics, Konan University
8-9-1 Okamoto, Higashinada, Kobe 658-8501, Japan
kunio@konan-u.ac.jp

Abstract. We developed a glasses-free 3D stereoscopic display using an LCD display panel and a special grating film for stereoscopic viewing. The display screen is divided in half in order that left and right regions provide the stereoscopic images for left and right eyes. Because both stereoscopic images are not in the same position, it is difficult for the observer to view the 3D image by the stereoviewing. The grating film can solve this problem because it shifts both left and right images to the same position. Moreover this grating film can give us glasses-free 3D viewing because of its view control effect. As the result, the each eye can perceive separated stereoscopic images for left and right eyes without special glasses such as polarized glasses.

Keywords: optical film sheet, grating film, viewing angle control, stereoscopic 3D imaging, display.

1 Introduction

Conventional 3D movie systems with the special glasses such as polarized glasses provide us touchable spatial images. However, these 3D imaging systems are very expensive and large scale equipment. Our research group would like to realize the simple 3D imaging system to construct an interactive spatial imaging environment. The authors have researched the 3D displays and applications. We have ever proposed 3D displays using the slit as a parallax barrier, the lenticular screen and the holographic optical elements(HOEs) for displaying active image. This paper describes the 3D display using an optical film, which is sold at D.I.Y. stores, for easy 3D image viewing.

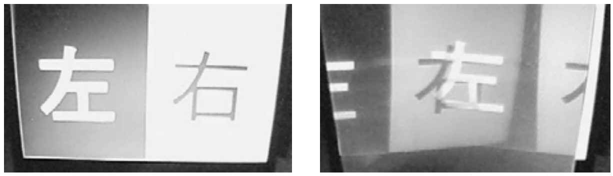
2 Superimpose Stereoscopic Images

To overlap left and right images, the authors use an optical film. This optical sheet is a flexible film with prisms designed to transport and diffuse the light. This sheet has interesting characteristics as follows; the prismatic phenomenon is observed and the doubling can be visible through the sheet like the Calcite. This doubling phenomenon occurs because the prism sheet diffracts two beams. These beams are called as the first order diffracted beam and the second order diffracted beam. Fig. 1 shows the doubling phenomenon. This interesting thing reminds us of method to superimpose left and right stereoscopic images. A grating diffracts or scatters a light beam with a designed angle. Using the doubling phenomenon, the authors shift the images for

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 449–451, 2010.

superimposing stereoscopic images by adjusting the interval between an optical sheet and image plane as shown in fig. 2. The optical grating film shifts both left and right images to the same position. The observer watches overlapped stereoscopic images for left and right eyes. This optical grating film is thin and works as a prism. The optical film appropriately locates apart from the display panel so that stereoscopic left and right images are overlapped. As shown in fig. 2(b), the left and right images are overlapped on the same plane. If these overlapped images can be separated into appropriate eyes, you can perceive the left image only by a left eye and the right image by a right eye.

Fig. 1. Doubling phenomenon



(a) side-by-side image (b) view thru grating film

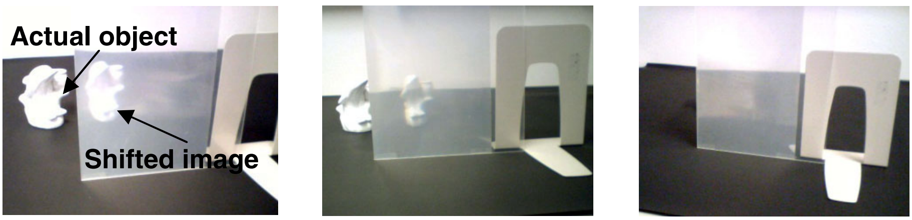
Fig. 2. Superimposing left and right images

3 View Control Effect of Grating Film

The authors firstly used the "SOLF™" optical film for overlapping left and right image as shown in fig. 2. But you need to wear polarized glasses for separating stereo image. To deliver left and right images into appropriate eyes though you wear no glasses, we use a new grating film with view control effect. The author unexpectedly found this sheet in D.I.Y. materials floor at Sannomiya Store of Tokyu Hands. This product name is unknown. We regard this sheet as a cheaper copy product of the SOLF sheet. However this sheet has a useful characteristic. Using this film, you can see through the film from the left, but not from the right as shown in fig 3. One of the miraculous features is that it can be either transparent or grating sheet, so that it looks either like transparent glass or prism, depending on the angle of sight.

It is useful characteristics for 3D viewing that you can control what can and what cannot be seen depending on which side the viewer is on, or what angle the viewer is looking from. Using the miracle of this visibility control as shown in fig 3, it enables us to perceive left images by the only left eye and right images by the only right eye. As shown in fig 4, the view control film passes the light within a designed angle. Let's design the optical layout assuming that 9-inch display panel is used. The width of the 9-inch panel is approximately 200mm. As shown in the fig. 4, the ray of a left image is emitted with an angle α to vertical and it reaches into the left eye after the ray is diffracted by an optical grating film. Meanwhile the grating film doesn't diffract the ray with an angle β, which passes into the right eye. Therefore the left image is observed by the only left eye because the ray with an angle β to vertical is blocked by the optical film. The rays of a right image are the same as the left image. Using this grating film, the observer can perceive the left image only by a left eye and the right image by a right eye with no glasses because the film with view control effect restricts the direction of scattering light after the grating film overlays left and right images at the same position. Therefore, the observers, who wear no glasses, can view the 3D images by the binocular viewing.

We have developed the prototype glasses-free stereoscopic 3D display using two commercial LCD panels for playing 3D contents by portable DVD players. The size of panels is 7.24-inch and its width of viewscreen is 103mm. Assuming the 9-inch panel, the interval between the center of viewscreen is 100mm. Both panels are side by-side and each LCD displays a left image for the left eye and a right image for the right eye. The grating film plate is positioned 92mm apart from the display panels so that stereoscopic left and right images are overlapped. The observer perceives an overlapped image through the grating film within a viewing window as shown in fig. 5. In trial display, the observation distance is 400 mm. Because of overlapping images by optical film, the observer can see the 3D image at approximately 300 mm apart from the viewing window of a display box. Since viewing positions of left and right images are restricted, the observers can view the 3D images by the binocular stereo viewing without special glasses.



(a) optical layout (b) viewing from left (c) viewing from right

Fig. 3. View control effect of grating film

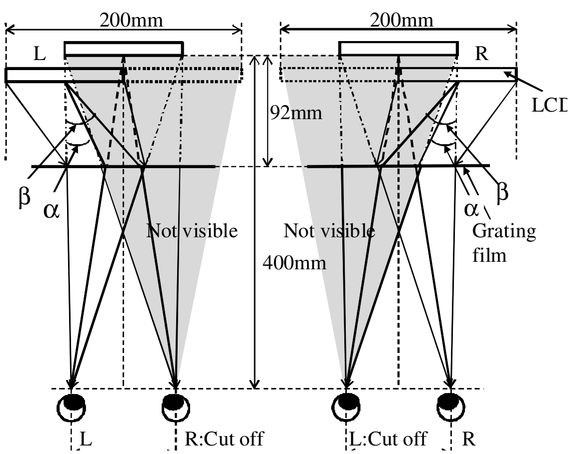


Fig. 4. Optical design for 3D viewing

Fig. 5. Appearance of 3D display (KNA-30)

Acknowledgments. This research is partially supported by "Grant-in-Aid for Young Scientists(B)" #20700112 and "Scientific Research (C) (General)" #20500481 from Ministry of Education, Culture, Sports, Science and Technology Japan(MEXT).

Omni-Directional Display System for Group Activity on Life Review Therapy

Tomoyuki Honda and Kunio Sakamoto

Department of Intelligence and Informatics, Konan University
8-9-1 Okamoto, Higashinada, Kobe 658-8501, Japan
kunio@konan-u.ac.jp

Abstract. The authors have researched support system of the reminiscence and life review activity. This support system consists of an interactive tabletop display and interface system. On the reminiscence and life review activity, a therapist puts pictures on the table so as to trigger a talk. However some observers may perceive upside down images if they sit down opposite the therapist. To overcome this problem, we have developed the display system which can be viewed from any direction. In this paper, we propose a 4-views tabletop flat display system for cooperative activity on a round table.

Keywords: all around viewing, group work, viewing angle control film, table-top display.

1 Introduction

The authors have researched multimedia system and support system for nursing studies on and practices of reminiscence therapy and life review therapy. The concept of the life review is presented by Butler in 1963. The process of thinking back on one's life and communicating about one's life to another person is called life review. A therapist must keep a record of sessions for inspection of methods and ways of valuation on reminiscence and life review therapy, but it is trouble for the therapist to record. The aim of research is to develop the support system which can automatically give an optimum topic and write down a session report about the activity. This life review is often assisted by aids such as videos, pictures, objects, archives and life story books, as shown in fig. 1, in order to make an opportunity of talking. We want to develop an omni-directional display system for cooperative activity on a round table to enable all-around viewing and unification of media contents by an electronic form.

2 Motivations

This paper describes an omni-directional display system that can be viewed from any direction (*i.e.*, the display has four viewing zones so as to perceive a screen view of the display at all directions around a table). The authors have ever researched information display systems involving 3D imaging. However, a conventional monitor

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 452–455, 2010.

display is viewed from one direction, that is, the display has narrow viewing angle and observers cannot view the screen from the opposite side. Hence we developed a 4-views display system for collaborative tasks cooperated by four users. This 4-views display can provide different images to four users surrounding the system utilizing an optical grating film for generating a virtual screen and floating above the top level of an actual display panel. The floating virtual screens of this display are generated into a square pyramid in front of observers' eyes. But this system displays different images in the pyramid and the display screen is not flat. Then this paper describes a flat type display system for the all-around viewing, which enables to put something on the table.

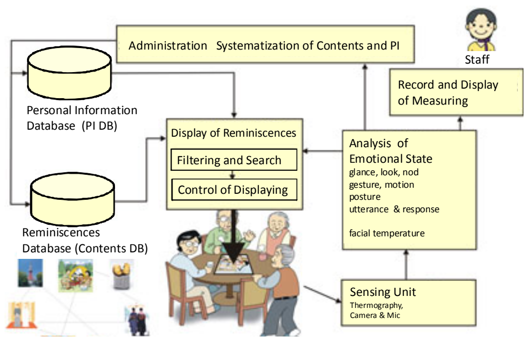


Fig. 1. Cooperative work of life review activity

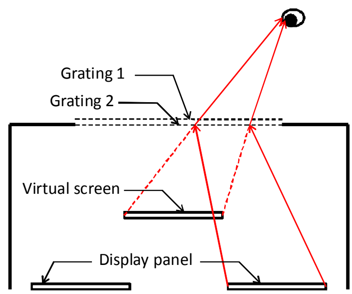
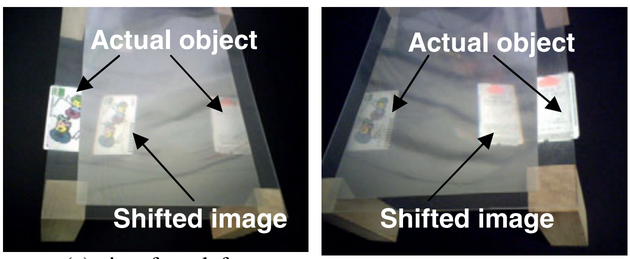


Fig. 2. Optical layout of display

3 4-Views Tabletop Flat Display

The authors firstly used the "SOLF™" optical grating sheet for shifting and floating image planes above the top level of an actual display panel. This display system is the shape where the outer surfaces are triangular and converge at a point. We would like to shape a new display like a conventional table. To enable all-around viewing from four directions, it is necessary to generate a virtual screen. So this viewscreen is floating above an actual display panel and sinking below the level of the table as shown in fig. 2. To simplify an optical layout, the authors utilize a special grating sheet for locating pseudo images at different places from original positions. The author unexpectedly found this special sheet at the Sannomiya store of Tokyu Hands. As shown in fig. 3, this grating sheet provides a diffracted image which is arranged left or right from original position. The grating sheet diffracts or scatters a light beam with a designed angle. This interesting phenomenon reminds us of method to shift image positions by a simple optical layout. This product name is unknown. We regard this sheet as a cheeper copy product of the SOLF sheet. However this sheet has a useful characteristic. Using this film, you can observe shifted image of the left card from the left, but not from the right as shown in fig 3. The right card is also observable as the same. Thus this new grating film has an effect of the viewing angle control in addition unlike an original SOLF sheet.

As shown in fig. 2, our developed omni directional display system consists of four LCD panels and two optical grating films for shifting original images into the center and separating virtual images into an appropriate eye. As above mentioned, one grating sheet can shift left and right images into the center and deliver each image to an observer. To generate four virtual screens, we piled up two grating sheets. The important point is that each sheet is inclined at 45 degrees to an edge of the table because both sheets shift the images into any direction independently. As shown in fig. 4, the combination of two optical sheets generates four directional image shifts. Each grating sheet shifts an actual image into any direction at 45 degrees to horizontal. Both image shifts makes vertical shift as the result. Other image shifts are also generated just the same. Fig 5 shows the appearance of our developed prototype display. Four display panels are arrayed crosswise in order to produce four virtual image screens into the center. A grating film appropriately locates apart from the panels so as to shift images into the center for delivering an appropriate image into each observer by adjusting the interval between the optical sheet and the display panels. The lights of an LCD pass through the films or are blocked by the films whether eyes are corresponding to each image or not. An observer can perceive delivered image without wearing special glasses.



(a) view from left (b) view from right

Fig. 3. Image shift by grating sheet

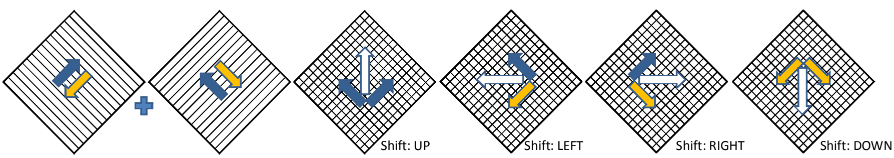


Fig. 4. Effect of image shift to four directions

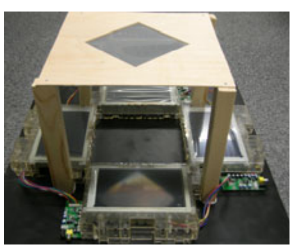


Fig. 5. Appearance of display (KNB-20)

We have developed the prototype display system using four commercial LCD panels. These displays have a 7.24 inch LCD panel and its size is 160mm(W) x 90mm(H). The video input supports NTSC. The panels are fixed on edges of a 220mm square. The size of a viewing window on surfaces of the table is 150mm square. This viewing window is with two grating sheets. The distance between these LCD panels and the viewing window is approximately 220mm so as to shift all images into the center. Since the grating sheet not only shifts an image but also controls a viewing angle, an observer perceives an only virtual screen right in front of his/her face. Observers can watch the screen of a display from any direction because this system has four screens with four viewing directions.

Acknowledgments. This research is partially supported by "Grant-in-Aid for Young Scientists(B)" #20700112 and "Scientific Research (C) (General)" #20500481 from Ministry of Education, Culture, Sports, Science and Technology Japan(MEXT).

Light-Weight Monocular 3D Display Unit Using Polypyrrole Linear Actuator

Yuuki Kodama and Kunio Sakamoto

Department of Intelligence and Informatics, Konan University
8-9-1 Okamoto, Higashinada, Kobe 658-8501, Japan
kunio@konan-u.ac.jp

Abstract. The human vision system has visual functions for viewing 3D images with a correct depth. These functions are called accommodation, vergence and binocular stereopsis. Most 3D display system utilizes binocular stereopsis. We have developed a monocular 3D vision system with accommodation mechanism, which is useful function for perceiving depth. This vision unit needs an image shift optics for generating monocular parallax images. But conventional image shift mechanism is heavy because of its linear actuator. To improve this problem, we developed a light-weight 3D vision unit for presenting monocular stereoscopic images using a polypyrrole linear actuator.

Keywords: head mounted display, monocular stereoscopic display, real-time stereogram, 3-D display.

1 Introduction

A study of virtual-reality system has been popular and its technology has been applied to medical engineering, educational engineering, a CAD/CAM system and so on. The 3D imaging display system has two types in the presentation method; one is a 3D display system using a special glasses and the other is the monitor system requiring no special glasses. A display system with no special glasses is useful for a 3D TV monitor, but this system has demerit such that the size of a monitor restricts the visual field for displaying images. The 3D display system using special glasses can display virtual images over a wide area. The human vision system has visual functions for viewing 3D images with a correct depth. These functions are called accommodation, vergence and binocular stereopsis. Accommodation is a useful function for perceiving a depth by the monocular vision system. Binocular stereopsis is one of the most important processes for the human 3D perception. Then most 3D display system utilizes this binocular stereopsis. The binocular vision mostly utilizes next two functions. One is simultaneous perception, which is an ability to perceive dichoptically presented images simultaneously and in the correct position. The other is binocular fusion, which is an ability to perceive two dichoptic images in left and right eyes as one image. Meanwhile accommodation enables us to perceive a depth by the monocular vision system. And humans have the function called the simultaneous perception. Using this function, humans perceive dichoptically presented images

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 456–459, 2010.

simultaneously and in the correct position. So the authors have developed a monocular 3D vision system which enables us to provide correct and natural 3D images with a depth perceiving by the accommodation.

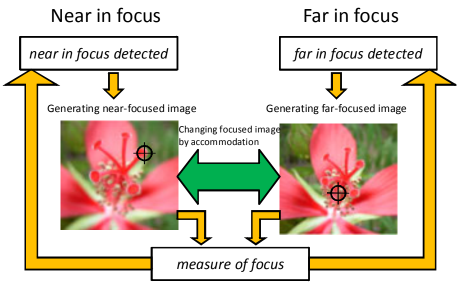


Fig. 1. Flow of generating focused image

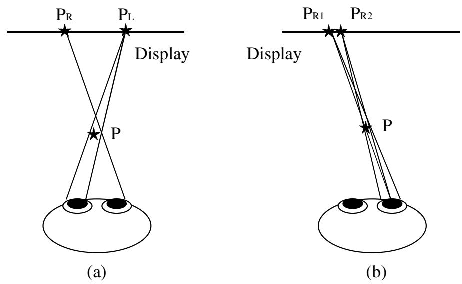


Fig. 2. The principle of a stereogram

2 Monocular 3D Display

To realize natural 3D viewing, we have developed the monocular vision system, which can directly project stereoscopic image on a retina, and a 3D image generation system, which can make 3D computer graphics in accordance with accommodation. Assume that an actual object is in the real world. When you perceive this object, a part of the projected image on a retina might be a blur by the lens of an eye. It is a physiological response called as accommodation. In case of virtual 3D image viewing as shown in fig. 1, you can watch a correct 3D image as the actual object is in there if the projected retina image has appropriate blur in compliance with focus adjustment of your eye. Then you might perceive virtual images with same accommodation as you watch real objects. Thus a monocular 3D vision system can provide correct 3D viewing with accommodation, vergence and binocular stereopsis and without a tired feeling at long time watching when the retina image is directly projected and external stimulation induces the focus adjustment by changing the thickness of an eye lens.

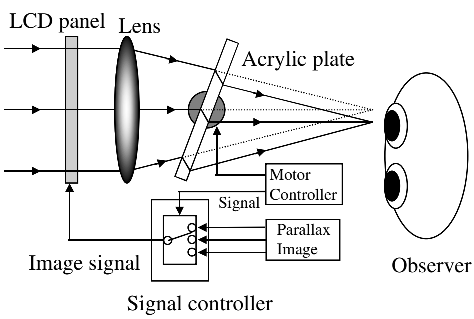


Fig. 3. Optical layout of monocular 3D display

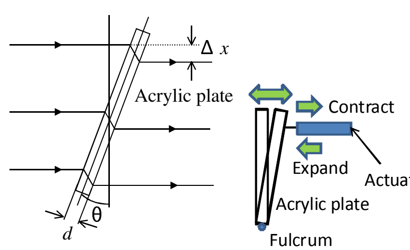


Fig. 4. Image shifting optics

Fig. 2 shows the principle of a reconstruction of the 3D image by the stereo-viewing. Fig. 2 (a) shows the optical configuration of a conventional stereogram. To display a point object P, observers turn on a point P_L for the left eye and a point P_R for the right eye as shown in this figure. The observer perceives that a point object exists on a spatial position P due to the binocular parallax. Fig. 2 (b) shows the optical configuration of a monocular multi-view stereogram. To display a point object P, observers turn on points P_{R1} and P_{R2} for the right eye as shown in this figure. At the monocular multi-viewing, the observer adjusts the focal length of an eye to match with the spatial position P, then projected images of pixels P_{R1} and P_{R2} are focused to the same position on the retina. So the observer naturally perceives that a point object exists on a spatial position P due to the monocular parallax. Fig. 3 shows the principle of the 3D vision system using monocular stereoscopy. This display system consists of an LCD panel, an acrylic plate and an optical lens. The observers perceive parallax images at the just point, which the optical lens converges the light on. To perceive multiple parallax images with just one eye, the image shifting optics consists of a parallel plane acrylic plate, whose inclination causes the image to shift as shown in fig. 3. An LCD panel is used as the displaying plane of parallax images. The signal controller sends an image signal to the LCD panel to the tune of a control signal. Then the displaying plane of parallax images creates monocular multi-viewing images.

3 Improvement of Image Shift Optics

As shown in fig. 4, the image shift Δx is generated by an inclined acrylic plate. To incline the plate, former vision unit used a motor. Therefore its weight is heavy so that observers wear it as glasses. To improve this problem, the authors developed new image shift mechanism for monocular 3D vision. This image shifting optics consists of an acrylic plate and a polypyrrole linear actuator as shown in fig. 4. Fig. 5 shows the structure of the polypyrrole actuator. This plastic film actuator is made of the polypyrrole. The thin layer of polypyrrole is attached on an acetate film as a support film. The polypyrrole can expand when it absorbs moisture in the air. When it discharges moisture, it can contract. Meanwhile the polypyrrole can also contract when the voltage is applied to it. The principle lies in the desorption of water vapor caused by Joule heating, where the electric field controls the sorption and the film absorbs and desorbs moisture reversibly in response to the applied voltage. The plastic film actuator is bent because the polypyrrole layer contacts but the acetate film cannot contract as shown in fig. 6. The polypyrrole film containing perchlorate is electrochemically synthesized by the anodic oxidation of pyrrole. The plastic film actuator can incline the acrylic plate when this film actuator is bent (or contracts) by controlling the applied voltage as shown in fig. 4. The film actuator is thin and light. So the authors could develop the light-weight monocular display unit for 3D viewing with a correct and natural depth using this polypyrrole actuator.

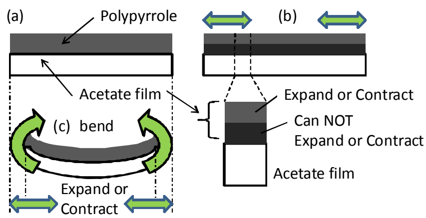
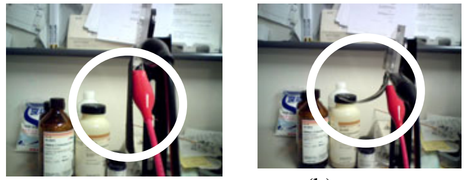


Fig. 5. Polypyrrole linear actuator



(a) off state (b) on state

Fig. 6. Appearance of polypyrrole linear actuator

Acknowledgments. This research is partially supported by "Grant-in-Aid for Young Scientists(B)" #20700112 and "Scientific Research (C) (General)" #20500481 from Ministry of Education, Culture, Sports, Science and Technology Japan(MEXT).

Easy Robot Programming for Beginners and Kids Using Command and Instruction Marker Card

Masahiro Nishiguchi and Kunio Sakamoto

Department of Intelligence and Informatics, Konan University
8-9-1 Okamoto, Higashinada, Kobe 658-8501, Japan
kunio@konan-u.ac.jp

Abstract. Robots usually have multiple components, such as motors, sensors, microcontrollers and embedded computers. A robot programming to control motors and measure the output of sensors is complicated. Therefore it is troublesome for beginners to write the program of a robot control. To solve this problem, this paper describes a card programming for controlling a robot.

Keywords: card programming, mobile robot control, structured programming.

1 Introduction

The authors have developed the mobile robot which can be programmed by command and instruction cards. All you have to do is to arrange cards on a table and to shot the programming stage by a camera. Our card programming system recognizes instruction cards and translates icon commands into the motor driver program. This card programming environment also provides low-level structure programming.

2 Instruction Sets for Robot Programming

Fig. 1 shows basic instruction sets to control a mobile robot. The mobile robot has wheels to move itself. ‘Go’, ‘Back’, ‘Left’ and ‘Right’ commands enables to transfer the robot anywhere you like. The wheels are rotated by motors. In order to move the robot, a motor driver needs to control the motors. Fig.2 shows the programming language for our developed motor driver and input/output controller. You must actually indicate which motor to work and how much it to rotate. However it is difficult for beginners to program and troublesome for all users to indicate in detail. So these basic commands are automatically translated into the program for the motor driver in our system. Moreover the authors developed the card programming environment using command and instruction marker which enables beginners or kids to enjoy lightheartedly the mobile robot operation. Fig. 3 shows examples of command and instruction cards.

Our developed motor driver can control six servo motors and have eight digital input/output ports. The motor driver has the PIC(Peripheral Interface Controller) microcomputer manufactured by Microchip Technology Inc. as a command service module and connects with a Windows PC through the RS-232C serial communication

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 460–463, 2010.

line. In our driver, the PIC microcomputer receives 26 kinds of instructions from the PC. One capital alphabet letter and an additional data constitute this instruction form as shown in Fig. 2. In order to control the mobile robot, the motor driver operates motors and turns a light on or off by the instruction, which is directly inputted or translated on the card programming environment.

Fig. 4 shows a situation of the card programming for our mobile robot control. As shown in Fig. 4, a camera shots the programming stage where you arrange the instruction cards on. On our card programming environment, you can instruct eight commands on the table at the same time. Moreover this card programming system supports a ‘for loop’ statement and a subroutine statement as the universal programming language. The ‘for loop’ allows command to be repeatedly executed. The subroutine also involves some commands to build a portion of instruction within a larger program. Then ‘for loop’ and subroutine have a series of instructions, and these statements behave in much the same way as a program that is used as one step in a larger program or another subprogram. The authors introduced and demonstrated the feasibility of a new concept in packaging. The packaging is to unify some instruction cards into one card as shown in Fig. 1. Thus packaging process can memorize and give a new function to a blank card which involves a series of commands and executes instructions step by step. Then to put a newly generated instruction card on the table means to call and execute the subroutine. The combination of the loop command and packaged instructions card realizes the loop function which executes same commands many times defined by a loop counter.

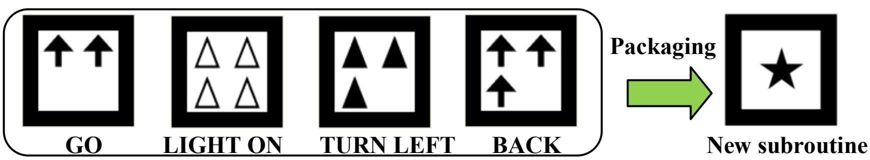


Fig. 1. Example of instructions and generating subroutine

GO	Move forward
BACK	Move behind
LEFT	Turn left
RIGHT	Turn light
LIGHT	Turn LED light on/off
LOOP	Execute instructions repeatedly
MEMORY	Package instructions (Generate subroutine)
CALL	Execute packaged instructions (Call subroutine)

```
10 U:01 70    #call subroutine
20 H:1        #Port H=high
30 W:100      #wait
40 H:0        #Port H=low
50 U:20 70    #loop 20 times
60 U:99       #program end
70 A:220      #Servo A=220
80 B:90       #Servo B=90
90 U:00       #return
```

Fig. 2. Example text programming

3 Easy Programming Using Command Cards

To realize the card programming, our developed easy robot programming system needs to recognize markers in the real world and discriminate command and

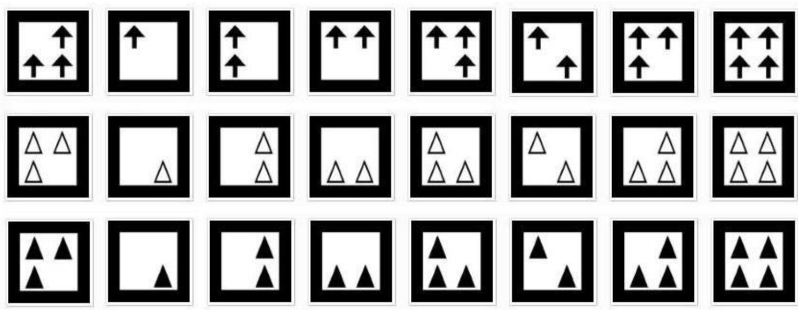
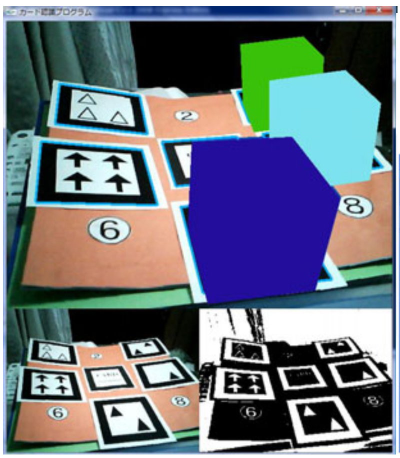


Fig. 3. Example of command and instruction cards



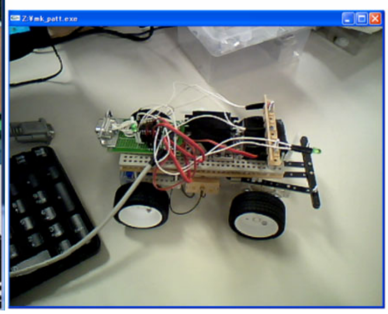


Fig. 4. Result of card recognition **Fig. 5.** Appearance of mobile robot

instruction from drawn patterns on the markers. As shown in Fig. 3, the marker card consists of a black square frame filled with some kind of a pattern. We shoot makers and take picture of it. The computer vision system firstly finds the black square flames from video streams and then recognizes patterns within the interesting regions. On our developed system, the authors make use of a mixed reality software in order to recognize a kind of the marker and measure its position. This software library is called ARToolkit. ARToolkit involves a video capture, a 3D graphics generator, a spatial measuring and an overlay imaging for the creation of augmented reality applications. Fig. 4 shows the result of card recognition. In Fig.4, the application window area consists of three regions; the upper screen shot shows the recognition result, the lower left shot is an original video stream and the lower right shows recognized instructions. The card programming stage has eight slots and the center is an origin marker for measuring card positions. A user puts the instruction cards onto these slots by turns of execution. At the example shown in Fig. 4, the slot number 2, 6 and 8 are vacant. The robot programming system finds the all instruction markers and overlays cubes on the markers as flags in case that the recognition process is correctly finished. As shown in lower right of Fig. 4, the six instructions are loaded excluding vacant slots. Fig. 5 shows the appearance of our mobile robot. You can control this robot to move anywhere you like and turn head lights on or off using the command and instruction cards.

Acknowledgments. This research is partially supported by “Grant-in-Aid for Young Scientists(B)” #20700112 and “Scientific Research (C) (General)” #20500481 from Ministry of Education, Culture, Sports, Science and Technology Japan(MEXT).

Automatic Mobile Robot Control and Indication Method Using Augmented Reality Technology

Koji Ohmori and Kunio Sakamoto

Department of Intelligence and Informatics, Konan University
8-9-1 Okamoto, Higashinada, Kobe 658-8501, Japan
kunio@konan-u.ac.jp

Abstract. A mobile robot is an automatic machine that is capable of movement in a given environment. Many techniques of automatic control are proposed. A line tracer is one of the most popular robots. The line tracer goes along a white line on the floor. The authors developed a mobile robot which moves to indicated point automatically. All you have to do is to indicate a goal point. In this paper, we propose an automatic mobile robot system controlled by a marker and remote indication using the augmented reality technology.

Keywords: mobile robot, automatic control, marker and remote indication.

1 Introduction

Our developed mobile robot can move itself to indicated goal point by an operator. A trial robot system provides us two indication methods. One is how to indicate as the operator put a marker as a goal point on the floor. The other is a remote control through closed-circuit television like the operator indicates the goal point on a video monitor. The mobile robot decides a route to the goal according to a behavior decision algorithm. The automatic mobile robot moves itself so as to shorten a distance from current residence to goal in consideration of a direction to the goal.

2 Target Indication by Goal Marker

Fig. 1 shows a scene of automatic mobile robot operation. The authors have developed a mobile robot system using a commercial radio control car and an additional controller. This robot is made utilizing a frame of the radio control car and attached with a marker in order to measure its position on the floor. The radio controller is also reconstructed so as to operate a motion from Windows PC. The operator indicates a goal using the marker as shown in Fig. 1. To control the mobile robot automatically, a camera fixed on the ceiling shots the markers which are robot and goal positions. The direction and distance from current residence to goal are calculated after the operation system recognizes markers and measures spatial positions. To get relation between markers, we used the software library which is called ARToolkit. ARToolkit involves a video capture, a 3D graphics generator, a spatial measuring and an overlay imaging

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 464–467, 2010.

for the creation of augmented reality applications. ARToolkit calculates three coordinate axes from a captured camera scene. The video tracking libraries calculate the real camera position and orientation relative to physical markers in real time as shown in Fig. 2.

Suppose that two 2D markers are laid on the floor. One is the marker which indicates a goal position. The rest shows the position of the mobile robot. The tracking library calculates each three coordinate axes on the two markers. Since our system program can detect a difference between two coordinate systems, we can know the direction and distance from current residence to goal. In Fig. 2, let (X_c, Y_c, Z_c) be three space coordinate of a camera and parameter Z_c shows the depth. Moreover, let (X_m, Y_m, Z_m) be three space coordinate of a marker and parameter Z_m shows the height. Assume that the transposed matrix $[X_c\ Y_c\ Z_c\ 1]^T$ means the camera coordinate and $[X_m\ Y_m\ Z_m\ 1]^T$ means the marker coordinate. Then the coordinate transformation between three space coordinates of a camera and a marker is given as follows by using a coordinate transformation matrix $\mathbf{T_{cm}}$;

$$[X_c\ Y_c\ Z_c\ 1]^T = \mathbf{T_{cm}}\ [X_m\ Y_m\ Z_m\ 1]^T. \quad (1)$$

Fig. 2 shows a transformation process between three space coordinates of markers on the same plane. Equation (1) shows the coordinate transformation between three space coordinate of a camera and three space coordinate of a marker. As shown in Fig. 2, the marker A and maker B are shot by the same camera. Assume that $[X_{Am}\ Y_{Am}\ Z_{Am}\ 1]^T$ means the coordinate of marker A and $[X_{Bm}\ Y_{Bm}\ Z_{Bm}\ 1]^T$ means the coordinate of marker B. From equation (1), the coordinate transformation matrixes $\mathbf{T_{Acm}}$ and $\mathbf{T_{Bcm}}$ are given as follows;

$$[X_c\ Y_c\ Z_c\ 1]^T = \mathbf{T_{Acm}}\ [X_{Am}\ Y_{Am}\ Z_{Am}\ 1]^T,\ \ [X_c\ Y_c\ Z_c\ 1]^T = \mathbf{T_{Bcm}}\ [X_{Bm}\ Y_{Bm}\ Z_{Bm}\ 1]^T.$$

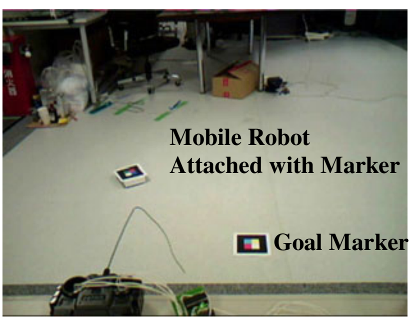


Fig. 1. Scene of mobile robot operation

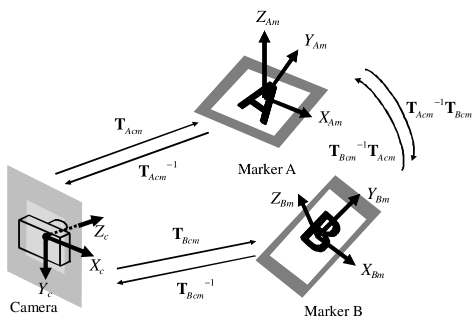


Fig. 2. 3D measuring for marker tracking

The coordinate transformation matrix $\mathbf{T_{cm}}$ shows the relation between three space coordinate of a camera and three space coordinate of a marker. When the coordinate transformation matrixes $\mathbf{T_{Acm}}$ and $\mathbf{T_{Bcm}}$ are determined, the camera is at the same position. Therefore the relation of the positions between the coordinate of marker A $[X_{Bm}\ Y_{Bm}\ Z_{Bm}\ 1]^T$ and the coordinate of marker B $[X_{Bm}\ Y_{Bm}\ Z_{Bm}\ 1]^T$ is given as follows;

$$[X_{Am}\ Y_{Am}\ Z_{Am}\ 1]^T = \mathbf{T_{Acm}}^{-1}\ \mathbf{T_{Bcm}}\ [X_{Bm}\ Y_{Bm}\ Z_{Bm}\ 1]^T,$$
$$[X_{Bm}\ Y_{Bm}\ Z_{Bm}\ 1]^T = \mathbf{T_{Bcm}}^{-1}\ \mathbf{T_{Acm}}\ [X_{Am}\ Y_{Am}\ Z_{Am}\ 1]^T.$$

After a decision of the direction and distance to the goal, the automatic mobile robot judges a next action itself according to an algorithm of the motion behavior decision. Fig. 3 shows the behavior decision algorithm. Suppose a line whose direction vector is the same to a direction of movement. Firstly the mobile robot judges whether a goal marker is in the left or right. Assume that the answer is the right. Next judgment is whether the goal exists in front or behind. This decision is to choice that the mobile robot goes straight or it makes a U-turn refers to performing a 180 degree rotation to reverse the direction of movement. If the goal exists in the same direction of movement, the robot goes straight or makes a U-turn in order to change a heading direction. In case that the goal does not lie on the line, the mobile robot moves to right in front or to left at back in order that it turns toward the goal. After slight movement the robot again judges whether a goal marker is in the left or right. To perform a series of slight movement reaches the goal. This robot can also take an action in the same way even if the goal marker exists in the left side.

3 Remote Control by Closed-Circuit Television

Fig. 4 shows the illustration of capturing the marker on the floor using a pin-hole camera. The marker in the real world and its perspective image on the plane of a film are indicated in this figure. This marker is utilized in order to define a plane of the floor. In Fig. 4, let (X, Y, Z) be three space coordinate and parameter Z shows the depth. The S_X-S_Y coordinate system shows the image plane of a camera. Assume three points and a normal vector. Points (p_x, p_y, p_z) and (t_x, t_y, t_z) are target points which are indicated on the video screen and in real world respectively. (m_x, m_y, m_z) is a center point of the square marker. (n_x, n_y, n_z) is a normal vector of the floor plane and it is also z axis of the marker coordinate. If the focal length of a camera is f , $p_z = f$. When you indicate point (p_x, p_y, p_z) on the view screen as a target point on the floor, points (p_x, p_y, p_z) and (t_x, t_y, t_z) lie on the same line. This line equation is

$$[t_x, t_y, t_z]^T = t\,[p_x, p_y, p_z]^T \quad (t \text{ is variable}). \tag{2}$$

Meanwhile the equation of the floor plane is given as follows;

$$n_x(X - m_x) + n_y(Y - m_y) + n_z(Z - m_z) = 0. \tag{3}$$

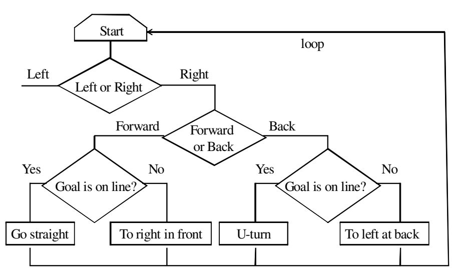


Fig. 3. Behavior decision algorithm

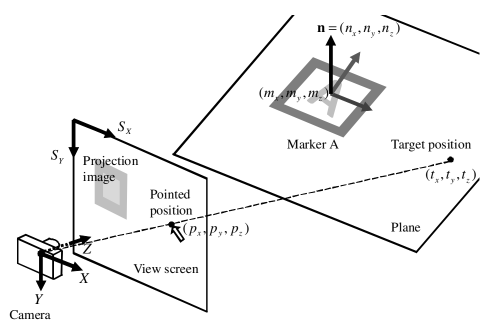


Fig. 4. Measurement of the 3D position

From the equations (2) and (3), we get an intersection (t_x, t_y, t_z) of the line and plane;

$$n_x(t\,p_x - m_x) + n_y(t\,p_y - m_y) + n_z(t\,p_z - m_z) = 0.$$

Then we can get the parameter t as follows and the indicated point on the real floor;

$$t = (n_x m_x + n_y m_y + n_z m_z) / (n_x p_x + n_y p_y + n_z p_z).$$

Using this parameter t, the mobile robot decides the goal point. After this calculation, the robot can move itself in the same way as the marker indication.

Acknowledgments. This research is partially supported by "Grant-in-Aid for Young Scientists(B)" #20700112 and "Scientific Research (C) (General)" #20500481 from Ministry of Education, Culture, Sports, Science and Technology Japan(MEXT).

Eye Contact Communication System between Mobile Robots Using Invisible Code Display

Takeru Furukawa and Kunio Sakamoto

Department of Intelligence and Informatics, Konan University
8-9-1 Okamoto, Higashinada, Kobe 658-8501, Japan
kunio@konan-u.ac.jp

Abstract. The authors have been developing the mobile robots which can cooperation between robots. The robots should communicate with each other in order to cooperate together. Therefore, the inter communication between robots is very important problem to be solved. These robots generally utilize wireless transmission system. The transmission sets send and receive on the same frequency or channel to establish the radio communication. This is called working simplex. The robots cannot start communication if both sets use different frequency channels. It is important to perform an initial configuration for establishing the radio signal transmission at a first contact among strange mobile robots. To solve this problem, this paper describes an information transmission system using an invisible code on displays which show an expression of robot's eyes.

Keywords: 2D code, polarized invisible code, polarized light control, LCD display.

1 Introduction

“The eyes are eloquent as the tongue.” This proverb means that one can say more with a look than with ten thousand words. In many cases, we perceive one's true intention from his or her facial expressions and gestures. This is a nonverbal communication. The nonverbal communications are usually understood as the process of communication through sending and receiving wordless messages. The eye contact is a basic and expected form of nonverbal communication as the famous quote by: “the eyes are the window of the heart, and the heart is the house for the soul.” This eye contact is a meeting of the eyes between two individuals. The authors hit upon an idea that fundamental transmission can make the best use of the eye communication. In this paper, the authors propose a communication method using an invisible code which is drawn on the eyes of the robot. We have developed the special graphics display for showing expression of robot's eyes and confirmed that our proposed technique enables the mobile robots to transmit and receive information data in order to establish a cooperation environment among robots.

A printing company has developed a new technology which makes a publication embedded with invisible codes, which is made by infrared reflective pigment, or small 2D codes, which is made by conventional offset printing using an infrared

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 468–471, 2010.

printing ink. The code symbol like a QR code consists of numerous intelligent small dots that can be read by a pen device. This pattern indicates the exact positions of the pen. When touching the pattern on the paper is automatically taken. Every snapshot contains enough data to determine the exact position of the pen device. Using this position data, the interactive computer system provides us with services such as a voice reading or a touch-panel-like operation on the paper. But this technology is realized only on the paper material. The authors would like to realize communication using an invisible code on an electrical display.

2 2D Code Contains Information Data

QR Code is a kind of 2D symbology developed by Denso Wave in 1994. QR Code (2D Code) contains information in both the vertical and horizontal directions, whereas a bar code contains data in one direction only. QR Code is capable of 360 degrees, high speed reading. QR Code accomplishes this task through position detection patterns located at the three corners of the symbol. This ability allows the use of 2-dimensional codes in a wide range of applications. QR codes onto labels are widely used in the fields of distribution and logistics. Moreover, QR Code is available for camera phones which enable new services based on QR-code (2D-barcode) input. The QR Code can store much information. The authors utilize this QR Code in order to establish the communication among the mobile robots.

(a) view of normal camera (b) thru horizontal polarizer (c) thru vertical polarizer

Fig. 1. Invisible display unit

3 Invisible Code Display

The printing technology using a special pigment enables us to provide a publication embedded with invisible codes. This technique is useful for developing an interaction system. We want to make good use of invisible codes at an electric display as well as a paper.

To display visual information and to embed invisible additional information, the display panel needs to hide code symbols so as not to interfere with screen viewing as shown Fig. 1. So we utilize a polarized symbol image to overlap additional information on the visual screen. The polarized light wave has a useful characteristic to generate hidden images. You know you cannot perceive digits of a calculator if a polarizer is removed from an LCD, *i.e.*, it is impossible for human's eyes to distinguish characteristics of polarization. In robot's eye expression display system

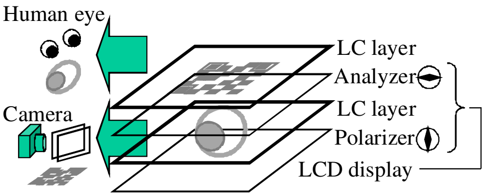


Fig. 2. Principle of generating invisible code

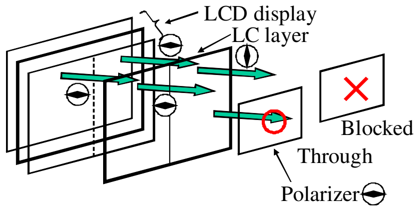
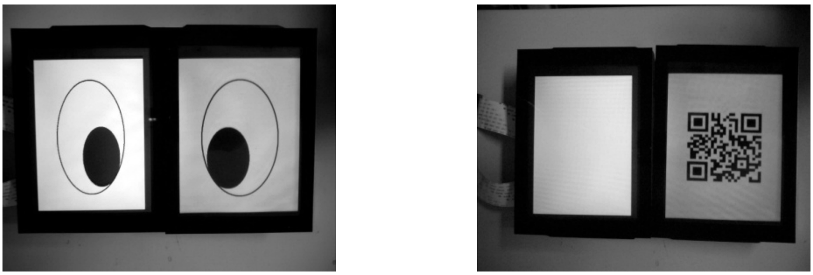


Fig. 3. Principle of hidden code detection



(a) expression of eye

(b) additional information contained in QR code

Fig. 4. Robot eye display

using LCD panels, we utilize characteristics of polarization. As shown in Fig. 2, our proposed display system consists of a conventional LCD panel, an additional liquid crystal (LC) layer and some optical elements. LC layers can rotate the direction of the polarization axis according to the applied voltage. The LC layer sandwiched between both polarizers displays visual information. This structure functions as an LCD panel. Then this LCD panel emits the polarized light due to the existence of a surface polarizer (it is called an analyzer). Moreover, the overlaid additional LC layer changes the direction of polarization from LCD outputs. This LC layer generates invisible symbol patterns. The final LC layer outputs horizontal or vertical linear polarized light waves. This difference of orientation makes a binary symbol image. As humans cannot perceive differences of polarization, they directly watch only visual images on the viewscreen without perceiving symbol patterns. At the detection, the polarized symbol pattern images are observable through the polarizer because the polarizer blocks the wave or not according to the orientation. This enables a camera to detect the invisible code on the display panel. Humans and cameras can perceive the hidden pattern through these optical elements. So the display panels show visual images and invisible symbols simultaneously. Human's eyes can get only visual information and a code reader finds an only binary symbol pattern. Fig. 3 clearly shows the basic concept of an invisible code display. The additional LC layer turns the linear polarization from LCD outputs. We can perceive the difference of polarization using a polarizer. Assume that the polarizer rotates 90 degrees. Then the direction of polarizer converts vertical into horizontal or vice versa. This means that a detector gets an inverse image which is white and black reversed. Using this invisible code, the display system provides humans with visual information like expressions of robot's eyes and additional information in order to perform an initial configuration for establishing the radio signal transmission at a first contact among strange mobile robots. Fig. 4 shows the appearance of the display panel for robot’s eyes. Fig. 4 (a) shows the observed image by a natural human’s eye. Fig. 4 (b) shows the detected image through the polarizer.

Acknowledgments. This research is partially supported by “Grant-in-Aid for Young Scientists(B)” #20700112 and “Scientific Research (C) (General)” #20500481 from Ministry of Education, Culture, Sports, Science and Technology Japan(MEXT).

The 'Interactive' of Interactive Storytelling: Customizing the Gaming Experience

Barbaros Bostan[1] and Tim Marsh[2]

[1] Interactive and Digital Media Institute (IDMI), National University of Singapore, Level 2, I-Cube Building No.21, Heng Mui Keng Terrace, 119613, Singapore
idmbb@nus.edu.sg
[2] Faculty of Arts & Social Sciences, National University of Singapore, Blk AS6, 11 Computing Drive, 117416, Singapore

Abstract. In this article, we define interactive storytelling as a gaming experience where the form and content of the game is customized in real time and tailored to the preferences and needs of the player to maximixe enjoyment. The primary focus of interactive storytelling should not be on the attributes of the technology or characteristics of the medium, such as the AI techniques, planning formalisms, story representations, etc. but on different interaction levels provided by computer games and basic components of player enjoyment such as difficulty levels and gaming rewards. In conducting an analysis of interactive storytelling systems, we propose a user-centered approach to interactive storytelling by defining different customization levels for an optimum gaming experience.

Keywords: Computer games, interactivity, interactive storytelling, gaming experience, user-centered design.

1 Interactive Storytelling: Beyond Story Structures

Interactive storytelling is a gaming experience where the form and content of the game is modified in real time and tailored to the preferences and needs of the player to provide a sense of control over the mutual discourse of play. After all, the fundamental goal of interactivity is to allow the user different choices to be able to receive a highly personalized end result [1]. From our point of view, player modelling or profiling is the first step in dynamic reconfiguration of the virtual world to provide personalized gaming experiences. And the second step is in the player-centered approach to interactive storytelling is to decide on the parameters to be adapted or reconfigured beyond story plots and artificial intelligence capabilities of virtual agents. The term reconfiguration is transforming the "push button" interactivity into the productive process of gameplay where the user makes significant interventions into a game world to produce dynamic effects throughout the system [2]. Given below is the figure showing possible customization levels for computer games and our proposed customization levels for interactive storytelling systems.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 472–475, 2010.

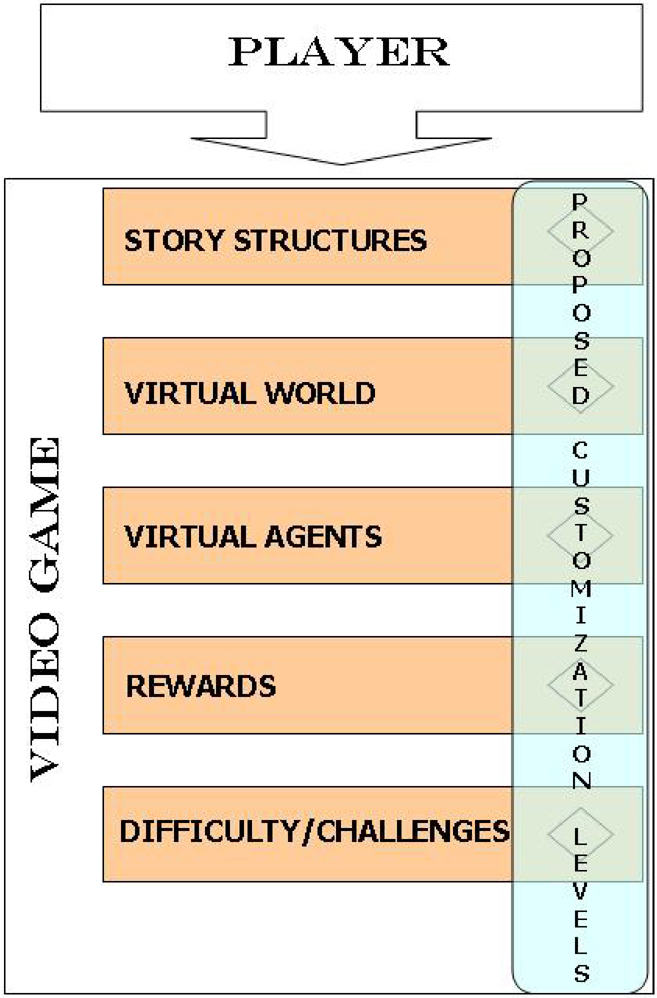


Fig. 1. Proposed customization levels for interactive storytelling systems

Customization of the gaming experience requires constant monitoring of the player actions within the virtual world. In this regard, Marsh et al. [3] proposed an approach to continuously and unobtrusively gather data about player interactions and a method to analyze and identify player behaviour, patterns and experience [4,5,6]. Player actions shall give an idea about his/her preferences or needs, but it should not be forgotten that the restrictions imposed by the game mechanics significantly reduce the number of player needs satisfied by a game. We claim that gaming experience should be customized for different interaction levels discussed in this study. Beyond story structures, the system should be capable of making real-time modifications on the virtual world (physical appearance, geography, regional history, life and society) and the virtual characters, as well as the player’s avatar (physical appearance, personality, relations with other characters, traits, motives). Two other customization levels for interactive storytelling systems are the rewards given (experience points, items, spells, titles, guild ranks) and the difficulty levels provided. Having defined these five customization levels (story structures, virtual world, virtual agents, rewards, difficulty), it is important to note that most of the commercial games focus on customizing story structures by player alignment (evil-neutral-good) and most of the research projects focus on customizing story structures by play styles. We claim that gaming experience should be customized on all these levels according to the preferences and needs of the player.

Although the majority of the interactive storytelling systems focus on the capabilities of the drama manager or the story planning processes, we believe that the "interactive" part of interactive storytelling is the user modelling or the player profiling that handles player interactions and defines how the experience should be customized for the player. Based on play styles, character classes or player types, the drama managers usually customize the story plots or NPC goals. But the critical questions are "what" to customize, "how" to customize and "why" to customize. Are story plots or character interrelations the only elements to be customized to provide an interactive gaming experience? This article already defined five different customization levels (story structures, virtual world, virtual agents, rewards, difficulty) for optimizing the gaming experience. And why is there a need for customization? In this regard, it should not be forgotten that motivation to reach a goal is influenced by both personal factors (needs, motivations, and goals) and situational factors (opportunities and possible incentives provided by the environment). Thus, we believe that gameplaying experience should be customized for maximizing player enjoyment or providing an optimum gaming experience which requires an understanding of player's psychology, especially his/her needs, motivations, and goals.

Are the play styles or player types in the literature the only way to portray player preferences? Although there have been attempts to define the basic motivations behind game playing, different approaches to motivational aspects of player behavior, when applied to computer games with different structures and content, represent different facets of player psychology and define several variables that leave much scope for subjective interpretation. In this regard, the motivational variables defined by the major studies in literature are more likely to shed light on the popular question "why do people play computer games," rather than predicting player behaviour or understanding player interactions/choices in a computer game. In order to define 'how' to customize the gaming experience, the motivational framework proposed by Bostan [7] should provide better insights since the study is based on psychological needs and analyzes these needs in relation to the gaming situations of a computer game. In an attempt to take this study one step further and to identify the common interaction patterns between these individual needs, Bostan and Kaplancali [8] applied the same motivational framework to another computer game and analyzed psychological needs by defining the driving game mechanics behind them. The same authors also analyzed user-created content (mods) of a popular computer game within the same motivational framework in terms of the needs they satisfy [9]. For future studies, this motivational framework should provide a convenient way of understanding player preferences.

References

1. Sundar, S.S.: Social psychology of interactivity in human-website interaction. In: Joinson, A.N., McKenna, K.Y.A., Postmes, T., Rieps, U.D. (eds.) The Oxford Handbook of Internet Psychology, pp. 89–102. Oxford University Press, Oxford (2007)
2. Dovey, J., Kennedy, H.W.: Game Cultures: Computer Games as New Media. Open University Press, Maidenhead & Milton Keynes (2006)
3. Marsh, T., Yang, K., Shahabi, C., Wong, W.L., Nocera, L., Carriazo, E., Kyriakakis, C.: Automating the Detection of Breaks in Continuous User Experience with Computer Games. In: Proceedings of CHI 2005 Human Factors in Computing Systems, Extended Abstracts, Portland, USA. ACM Press, New York (2005)
4. Yang, K., Marsh, T., Mun, M., Shahabi, C.: Continuous Archival and Analysis of User Data in Virtual and Immersive Game Environments. In: The 2nd ACM Workshop on Capture, Archival and Retrieval of Personal Experiences (CARPE 2005), Singapore (2005)
5. Marsh, T., Yang, K., Shahabi, C.: Game development for experience through staying there. In: ACM SIGGRAPH Video Games Symposium, Boston, USA. ACM Press, New York (2006)
6. Marsh, T., Smith, S., Yang, K., Shahabi, C.: Continuous and Unobtrusive Capture of User-Player Behaviour and Experience to Assess and Inform Game Design and Development. In: 1st World Conference for Fun ’n Games, UK (2006)
7. Bostan, B.: Player Motivations: A Psychological Perspective. ACM Computers in Entertainment 7(2) (2009)
8. Bostan, B., Kaplancali, U.: Explorations in Player Motivations: Game Mechanics. In: Proceedings of GAMEON 2009, Düsseldorf, Germany (2009)
9. Bostan, B., Kaplancali, U.: Explorations in Player Motivations: Game Mods. In: Proceedings of GAMEON-ASIA 2010, Shanghai, China (2010)

Remote Context Monitoring of Actions and Behaviors in a Location through the Usage of 3D Visualization in Real-Time

John Conomikes[1], Zachary Pacheco[1], Salvador Barrera[2], Juan Antonio Cantu[2], Lucy Beatriz Gomez[2], Christian de los Reyes[2], and Juan Manuel Mendez Villarreal[2]
in collaboration with
Takao Shime[3], Yuki Kamiya[3], Hideki Kawai[3], Kazuo Kunieda[3], and Keiji Yamada[3]

[1] Carnegie Mellon University, Entertainment Technology Center (ETC), 800 Technology Drive, Pittsburgh, PA, 15219, USA
[2] Universidad de Monterrey (UDEM), Engineering and Technology Division, Av. Morones Prieto 4500 Pte. San Pedro Garza Garcia, C. P. 66238, N.L. Mexico
[3] NEC C&C Innovation Research Laboratories, 8916-47, Takayama-Cho, Ikoma, Nara 630-0101, Japan
{jconomik,zpacheco}@andrew.cmu.edu,
{sbarrer1,jcantaya,lgomez20,xpiotiav,jmndezvi}@udem.net,
t-shime@ce.jp.nec.com, y-kamiya@fn.jp.nec.com,
h-kawai@ab.jp.nec.com, k-kunieda@ak.jp.nec.com,
kg-yamada@cp.jp.nec.com

Abstract. "Remote Context Monitoring of Actions and Behavior in a Location Through the Usage of a 3D Visualization in Real-time" is a software application designed to read large amounts of data from a database and use that data to recreate the context that events occurred to improve understanding of the data.

Keywords: Multi3D, Visualization, Remote, Monitoring, Panda3D, Real-Time.

1 Introduction

This prototype is the result of a long project development made at the Entertainment Technology Center where work was done in conjunction with NEC and the Universidad de Monterrey.

The goal of this project is to present huge amounts of data, not parse-able by a single person and present it in an interactive 3D recreation of the events that the sensors detected using a 3D rendering engine known as Panda3D.

While there is a lot of work in this field one of the unique angles of this project is the type of data is designed to build the recreation from. This data is from NEC's LifeLog system which tracks a wide variety of detailed information on what each employee in the monitored space does daily on a second to second basis.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 476–478, 2010.

Additionally, the data can be viewed from anywhere in the world, not just the monitored laboratory.

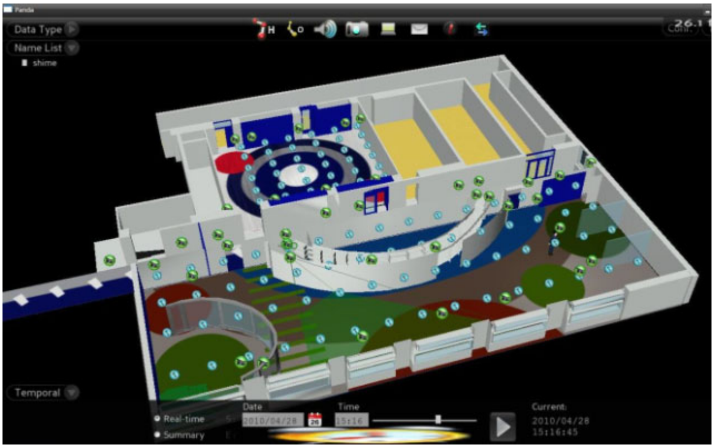


Fig. 1. Initial 3D shaded model for the Southern laboratory

2 System Architecture

Our entire system is built on NEC's LifeLog system which is responsible for gathering the large amount of data that is needed for the software to operate. See Figure 2 below for a view of the ceiling with installed sensors.

Fig. 2. Ceiling of the South Laboratory with installed sensors

Employee location is detected through the use of IR emitters on employees and receivers mounted on the ceiling, though approximately 25% of all location data is "8022" which is the code for a person who is not detected by any IR receiver on the premises. Ambient sound level data is collected by over 90 microphones installed in the ceiling.

All E-mails send to or from monitored employees are also stored, though addressees that are not monitored are stored only as “Company Employee" or “Recipient Outside Company".

Additionally, extensive information is pulled from the computer operations of each monitored employee. Statistics such as key presses, mouse clicks and mouse movements in the past second. Further, they track the currently active process running on the computer and the most recently accessed file. Even all of the currently running processes in the background. Finally they log all of the employee's internet access, though this last piece of information can be disabled by the employee. Finally, each employee has a wireless button that they carry with them that records when it was pressed and if pressed for more than one second, it also reports the duration of the press.

Also, while not related to people, 16 RFID readers are used to track the location of resources (e.g. books, laptops) which have RFID tags on them, as they move around the office. It also tracks which employee is using each particular resource.

The flow of information is quite simple, the LifeLog system polls the sensors for their latest information. It then takes this information, timestamps it and outputs it to a simplified YAML[1] format and stores this information on a server. Out program then connects to the server and requests the files required to view the time the user wishes to view, loads the needed information into memory in python data structures and displays the recreated events to the user.

Due to security restrictions at NEC, the data is only accessible locally or through a Virtual Private Network (VPN) connection. However, since the only remote action that is being performed with the software is reading data from the server, with less strict security measures, the software can function anywhere without the need for any special access permissions.

3 Experimental Results and Conclusion

While the system does use a large concentration of sensors in a small area and is generally very invasive, it does mean there are many promising opportunities for future research to improve on both the technology and software. While not ready for industry yet, with the inclusion of other research as well further improvement of the current software this seems to be a promising technology and may prove to be the next big step in combining multiple different information gathering technologies.

References

1. Hibbard, B.: Visad: connecting people to computations and people to people. SIGGRAPH Computer. Graphics. 32(3), 10–12 (1998)

Wave Touch: Educational Game on Interactive Tabletop with Water Simulation

JoongHo Lee[1], Won Moon[1], Kiwon Yeom[1], DongWook Yoon[1], Dong-young Kim[2], JungHyun Han[2], and Ji-Hyung Park[1]

[1] Intelligence and Interaction Center, Korea Institute of Science and Technology, Korea
{yap153,moonwon,pragman,dwyoon,jhpark}@kist.re.kr
[2] Game Research Center, College of Information and Communication, Korea University
{sadiles,jhan}@korea.ac.kr

Abstract. In this paper, we present an underwater exploration game called Wave Touch, designed specifically for a category of devices known as, interactive tabletops. The game provides users with a fun way to learn about important historical artifacts. An emphasis is placed on making Wave Touch entertaining to the user, a goal which is satisfied through the use of interactive tabletops and realistic water simulation. We also present the techniques we used to enable real-time water simulation effects in the game.

Keywords: tabletop, water simulation, educational entertainment, multi-touch.

1 Introduction

As a way of demonstrating the powerful potential of interactive tabletops for the entertainment domain, we developed an educational game called Wave Touch. Briefly, Wave Touch teaches players about significant historical artifacts pertaining to Korean culture. We also incorporated realistic water simulation effects to provide an exciting dynamic to the game environment. This captivating visual feedback mechanism can also serve to increase user enjoyment while playing Wave Touch.

Many researchers have dealt with tabletop technologies for entertainment purposes. In [1], Kaltenbrunner et al. presented reacTable, a tangible interface on an interactive tabletop, for musical performances. Tse et al. demonstrated a multimodal interface using a combination of hand gestures and verbal utterances for playing games on tabletops [2]. In [3], Gross et al. designed and evaluated a multi-touch tabletop interface, which supports cooperative and competitive gaming.

These days, a few researchers have shown interest in real-time fluid simulation. Bridson et al. [4] presents the curl-noise method for procedural fluid flow and added interesting fluid motions by using simple vector calculus. Thurey et al. [5] presents a new method for enhancing water simulations using overturning waves. This method can simulate large bodies of water in real-time. Yuksel et al. [6] introduces a novel method for real-time simulation of fluid surface waves. This method is very simple, fast, and unconditionally stable. It can simulate large bodies of water in real-time.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 479–481, 2010.

2 Interactive Tabletop

One of the most crucial elements for interactive tabletops is the touch sensing technology, particularly multi-touch, because they usually don't provide any peripheral devices for interaction (e.g. mice, keyboards). Interactions involving touch gestures tend to provides for a visceral user experience. Users feel much more engaged with the interface and derive more enjoyment out of the experience than with other conventional devices. There are a myriad of possible approaches for multi-touch detection, the two more well known approaches are capacitive [7] and optical-based [8]. We have developed novel optical multi-touch sensor to apply to Wave Touch.

3 Water Simulation

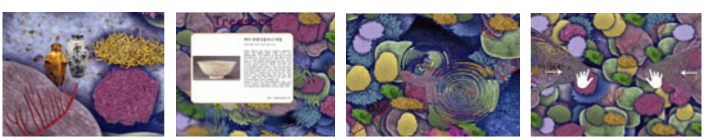
The Saint-Venant equations are a generalized version of the N-S equations. These equations reduce the problem of a 3D fluid motion to a 2D description with a height-field representation. The conservative Saint-Venant equations can be written as

$$\frac{\partial h}{\partial t} + \frac{\partial hu}{\partial x} + \frac{\partial hv}{\partial y} + \nabla \cdot (hU) = 0 \qquad \frac{\partial hu}{\partial t} + \frac{\partial hu^2}{\partial x} + \frac{\partial huv}{\partial y} = -\frac{\partial}{\partial x}\left(g\frac{h^2}{2}\right) \qquad \frac{\partial hv}{\partial t} + \frac{\partial huv}{\partial x} + \frac{\partial hv^2}{\partial y} = -\frac{\partial}{\partial y}\left(g\frac{h^2}{2}\right)$$

Where h is the height of the fluid above zero-level, U is the velocity of the fluid in the horizontal plane, u and v is each component of the velocity of the fluid U, and g is gravity. The Saint-Venant equations are commonly solved using the CPU, but solving the Saint-Venant equations on the CPU is very slow and not suitable for real-time applications. Therefore, we solve these equations using graphics hardware instead of the CPU, which significantly improves simulation performance and allows simulation to be performed on high resolution grids in real-time.

4 Wave Touch Game

The main goal of Wave Touch is to search and find hidden treasures (Fig. 1a) that are scattered throughout the underwater environment. As users navigate throughout the environment, these treasures will appear and disappear randomly. When a user finds and selects a treasure item, the game will pause momentarily and an information panel (Fig. 1b) will appear about that treasure. Each treasure found will award the user with a certain number of points. The game ends when all treasures have been found, and the player with the most points will win.



(a) (b) (c) (d)

Fig. 1. (a) Treasures (b) Information Panel (c) Multi-touch Interaction (d) Zoom-out gesture

Since Wave Touch is played on an interactive tabletop, players play the game by directly touching the tabletop interface. Navigation throughout the game environment is accomplished by using various single touch and multi touch gestures. The game world is presented in 2 dimensions so navigation consists of panning an overhead point-of-view or zooming in or out. To shift the point-of-view location, the player needs to swipe a finger across the interface. The scene will then move in that direction. To zoom in or out of the environment, the player needs to use a multi-touch pinch gesture, which entails bringing two contact points apart or together (Fig. 1d), respectively. When a player locates a treasure, they should directly touch that treasure on the screen in order to capture it.

5 Conclusion

We explored interactive tabletop technology and water simulation effects, particularly in regard to their contributive effects for entertainment related applications. A Series of user evaluation show that interactive tabletops are powerful devices for the entertainment domain due to the unique interaction modalities that are possible and the social nature of the device. Additionally, realistic water simulation effects provide applications with a type of visual feedback that is exciting and captivating, which can also improve user enjoyment.

References

1. Kaltenbrunner, M., Jorda, S., Geiger, G., Alonso, M.: The reactable*: A collaborative musical instrument. In: WETICE 2006: Proceedings of the 15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, Washington, DC, USA, pp. 406–411. IEEE Computer Society, Los Alamitos (2006)
2. Tse, E., Greenberg, S., Shen, C., Forlines, C.: Multimodal Multiplayer Tabletop Gaming. In: PerGames 2006: Proceedings Third International Workshop on Pervasive Gaming Applications, in Conjunction with 4th Intl. Conference on Pervasive Computing, pp. 139–148 (2006)
3. Gross, T., Fetter, M., Liebsch, S.: The cuetable: cooperative and competitive multi-touch interaction on a tabletop. In: CHI 2008 Extended Abstracts on Human Factors in Computing Systems, pp. 3465–3470. ACM, New York (2008)
4. Bridson R., Hourihan J., Nordenstam M.: Curl-noise for procedural fluid flow. ACM Transactions on Graphics (In Proc. of ACM SIGGRAPH) (2007)
5. Thurey, N., Muller-Fischer, M., Schirm, S., Gross, M.: Real-time breaking waves for shallow water simulations. In: Proc. of the 15th Pacific Conference on Computer Graphics and Applications, pp. 39–46 (2007)
6. Yuksel, C., House, D. H., Keyser, J.: Wave particles. ACM Transactions on Graphics (In Proceedings of ACM SIGGRAPH) (2007)
7. Dietz, P., Leigh, D.: Diamondtouch: A multi-user touch technology. In: Proceedings of UIST 2001, pp. 219–226. ACM Press, New York (2001)
8. Han, J.: Low-cost multi-touch sensing through frustrated total internal reflection. In: Proceedings of UIST 2005, pp. 115–118. ACM Press, New York (2005)

Healthy Super Mario: Tap Your Stomach and Shout for Active Healthcare

Jaewook Jung, Shin Kang, Hanbit Park, and Minsoo Hahn

Digital Media Laboratory, KAIST,
517-10 Dogok-dong, Gangnam-gu, Seoul 135-854, S. Korea
{jaewookjung,kangshin12,han_bit,mshahn2}@kaist.ac.kr

Abstract. The purpose of game provides fun and enjoyment to users. Most of the game, however, has physical dysfunction with providing fun experience. As a solution about this problem, the body movement based game like Nintendo Wii was presented. However, it just prevents the physical dysfunction by adopting body movement as game input and requires the special controller to play a game. In this research, we suggest a new game input style which is tapping stomach and shouting. It is not only preventing the dysfunction but also promoting health.

Keywords: Game interface, Interaction style, Active healthcare.

1 Introduction

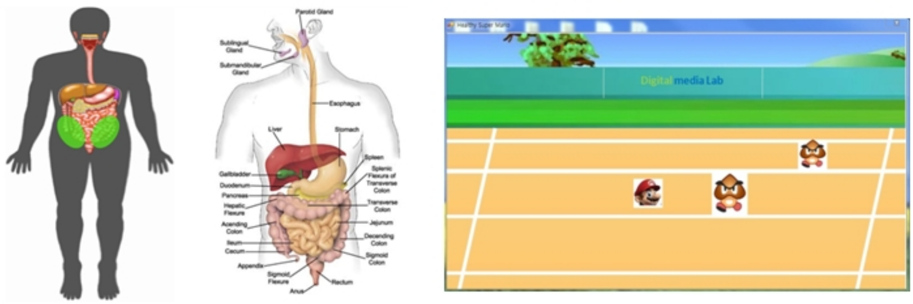
A physical dysfunction has been occurred with the growth of game industry [1]. As a physical dysfunction by game, there are Visual Display Terminal (VDT) syndrome, Spinal curvature, and Carpal Tunnel syndrome. To solve the problems, the exertion game interface like Nintendo Wii or Dance Dance Revolution [DDR] was developed [2], but it just prevented the physical dysfunction. We suggest two new game input methods for active healthcare beyond just preventing ill health. One is tapping stomach with both hands. Some medical research was shown that tapping stomach by both hands promotes the secernent of digestive fluid and the circulation of the blood. In other words, it promotes a digestive process [3]. The other style is shouting like a kendo player. Shouting like a kendo or taekwondo player is effective in reducing stress and improving concentration [4]. We suggest a new game interface that detects tapping the stomach by applying digital image processing (DIP) technology and measures the shouting level by digital signal processing (DSP) technology. We also present new athletic game, *Health Super Mario*, with the suggested interface.

2 Game Interface and Healthy Super Mario

The suggested interface consists of one webcam and one normal PC. The webcam detects the tapping stomach by both hands and counts the number of times tapping the stomach occurs in interval of 500 ms. The green colored ellipses in Figure 1-(a) is

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 482–483, 2010.

the tapping area and intestines are located in Figure 1-(b). The microphone of the webcam detects shouting by a player and measures the volume of shouting.



(a) (b) (c)

Fig. 1. (a) An anatomical chart, (b) A digestive system, and (c) The screenshot of *Healthy Super Mario*

Figure 1-(c) shows that *Health Super Mario* athletic game played by the suggested interface. Mario runs and jumps to avoid *Goomba*, an enemy character. The running speed is proportional to the number of tapping stomach per unit time and Mario jumps by shouting.

3 Conclusion

Many office workers or youth immediately play an online game or web surfing after eating a meal. This behavior has a bad influence on a digestive process. In this study, we suggest new input style that uses tapping stomach and shouting as game input. It aims to prevent the physical dysfunction and builds up health by changing a game style in life. As a result, the suggested input style is an entertainment system for active health.

References

1. Jung, J., Hahn, H., Yamazaki, T.: Game Interface Design Guidelines to Solve the Dysfunction of an Online Game. In: 21st International Symposium Human Factors in Telecommunication, pp. 359–366. Prentice Hall, Englewood Cliffs (2008)
2. Wii, N.: `http://www.nintendo.com/wii`
3. Kim, J., Park, J.: The Effect of Abdominal Meridian Massage on the Degree of Constipation among the Patients with Immobility after Orthopedic Surgery. Clinical Nursing Research 12(2), 63–72 (2006)
4. Kang, K., Yun, J.: The Effects of shouting on Motor Neural Adaptation and Maximum Isometric Contraction. Korean Journal of Sport Science 12(4), 85–93 (2001)

Interactive Manipulation Model of Group of Individual Bodies for VR Cooking System

Atsushi Morii, Daisuke Yamamoto, and Kenji Funahashi

Nagoya Institute of Technology,
Cokiso-cho, Showa-ku, Nagoya, 466-8555, Japan
morii@center.nitech.ac.jp,
yamamoto.daisuke@nitech.ac.jp,
kenji@nitech.ac.jp
http://www.center.nitech.ac.jp/

Abstract. A new high-speed manipulation model for a group of individual bodies (GIB) such as sand and lava is proposed in this paper. One of the goals of this research is to use it for applications such as home VR cooking systems for representing eg. a mass of fried rice. In this model, GIB is represented as a height field. Variation in the height field represents movement of the GIB. Transformation of GIB in wide spaces which is beyond adjacent grids is considered. GIB is treated as one object, which means that calculation is done efficiently on one object. Transformation of GIB is calculated quickly. Interactivity has a priority over correct movement of GIB in this model.

Keywords: a group of individual bodies, VR cooking system, home use.

1 Introduction

In recent years, research about interactive manipulation for sand and lava has been performed [1] [2]. In these researches, transformation of a group of individual bodies (GIB) such as lava and sand in a wide spaces which is beyond adjacent grids needs multiple calculations over a timeframe. For example, in a sand manipulation model [2], the collapse of the GIB is calculated by searching adjacent grids on every timeframe. The transformation of the GIB surface is conducted by sequences of calculations. In this paper, we propose a high-speed GIB model which only needs a few calculations for a certain point of time. In this model, we consider transformation of GIB in wide spaces which is beyond adjacent grids. GIB is treated as one object, which is calculated quickly.

2 Manipulation Model of GIB

Transformation Surface. In this model, GIB is defined as a height field which is fixed on a container. Transformation of the GIB is calculated from a curved surface. When the GIB exists in the center of a container, for example (Fig.1-1), transformation surface (TS) is generated around place where the GIB seem

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 484–486, 2010.

to slide down to (Fig.1-2 the black part), and the volume is added to the GIB (Fig.1-3). Then the GIB volume is corrected (Fig.1-4), and as a result the GIB is moved. A TS is generated as a half of elliptic cylinder for simplified calculation [3]. For example, when the container is tilted, a TS is generated in Fig.2. When an acceleration is applied to the container, the GIB receives a reaction force. The TS is calculated from the gravity and the reaction force.

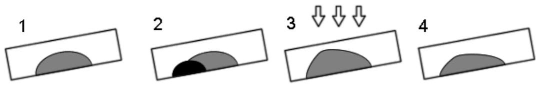


Fig. 1. Vertical Section of GIB and container

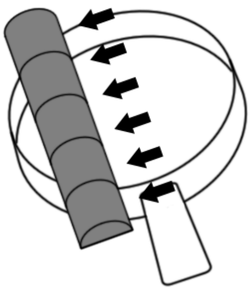


Fig. 2. TS of a half of elliptic cylinder

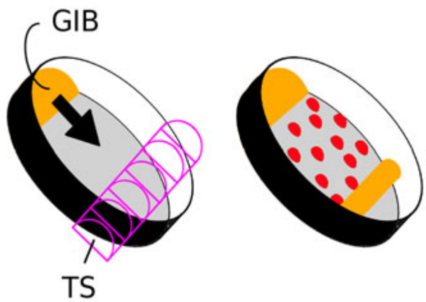


Fig. 3. Process of GIB movement

Expression Based on Existence Probability. Though TS model is quick because of its simple method, this model is insufficient in expressing the GIB's movement. To represent the process of GIB movement, we conduct an expression based on existence probability. When TS is generated on the place where TS is isolated from GIB in Fig.3 (left), we suppose that GIB are sliding down on the container. In this case, the GIB spreads both at the top of container and TS. Each grid of height field in the container are added random values based on existence probability (Fig.3 right). This expression is continued as long as more than a threshold volume exists between the center and the top of the container.

Particle System. We also implement a particle system [4] in this model. In order to implement the system, we locate the particles in height field values. Particle collision is detected on the surface of the height field. They also fall down when the container is tilted, and then the total volume of the GIB in the container is decreased. Only a few particles are used for maintaining a level of realism.

3 Experiment and Conclusion

The experiment was constructed on a system with a WiiRemote using a Intel Core 2 Duo E6300 1.86GHz CPU, 1024MB RAM. In this system, the GIB is

Fig. 4. System Appearance

supposed as fried rice (not traditional Chinese fried rice but Japanese) and the container as a skillet (Fig.4). Through this system, subjects were able to move fried rice by moving the container. 8 subjects evaluated the GIB movement's realism by classifying the score into seven grades. 1st grade is "This model doesn't feel real enough for a cooking system". 7th grade is "It is as natural as a real one". 3 subjects evaluated it as 4. 4 subjects evaluated it as 5. One subject evaluated it as 6. The results of the performance test is shown in the Table 1. We got some positive evaluations such as that GIB movement seems natural with 331 vertices and 670 particles. Good results were obtained but there are some points that need improvement. We have to improve upward movement of the GIB without the particle system. We are also developing a model for using cooking tools e.g. a spatula in this model.

Table 1. Performance Results

Vertices	Particles	FPS
331	670	20
631	1300	18
817	1700	15

References

1. Stora, D., Agliati, P.O., Cani, M.P., Neyret, F., Gascuel, J.D.: Animating lava flows. In: Proc. of the Graphics Interface, pp. 203–210 (1999)
2. Onoue, K., Nishita, T.: Virtual Sandbox. In: Proc. of the 11th Pacific Graphics, pp. 252–259 (2003)
3. Morii, A., Yamamoto, D., Funahashi, K.: Grid and particle based model of group of individual bodies for VR cooking system. In: Proc. of the 14th Virtual Reality Society of Japan Annual Conference (DVD) (2009) (in Japanese)
4. Harada, T., Tanaka, M., Koshizuka, S., Kawaguchi, Y.: Real-time Coupling of Fluids and Rigid Bodies. In: Proc. of the APCOM 2007-EPMESC XI (2007)

Smile Like Hollywood Star: Face Components as Game Input

Jaewook Jung, Hanbit Park, Shin Kang, and Minsoo Hahn

Digital Media Laboratory, KAIST,
517-10 Dogok-dong, Gangnam-gu, Seoul 135-854, S. Korea
{jaewookjung,han_bit,kangshin12,mshahn2}@kaist.ac.kr

Abstract. Most of the commercialized games have used controller by manipulated hands or foots. Recently, some studies tried to facial expressions or emotion as game input, but it didn't directly manipulate a game. It adjusted the difficulty of game. In this study, we suggest a new type of game input interface using face components as input, directly, and present one offline game and two online games by using this method.

Keywords: Game interface, Face components, Smile.

1 Introduction

Many offline games use joystick as game input and most of the online games use a keyboard and a mouse as game input. Some researches tried to adopt facial expression as game input indirectly [1-3]. Those studies measured the emotional state from the facial expressions and adjusted the level of difficulty in game. It is not the direct input for game.

In this study, we present the face component recognition (FCR) interface. In addition to the interface, we suggest one offline game, *Smile like Hollywood Star,* controlled by FCR, and apply FCR to two online games, Tetris and Kart-Rider.

2 Face Component Recognition Interface

FCR interface consists of one webcam, one normal PC, and one keyboard signal generation board. Most of the online games require the keyboard signal in hardware level because of security. Therefore, we implemented the signal generation board to occur the real key-code signal in Figure 1.

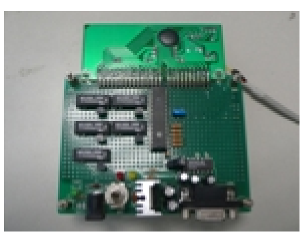


Fig. 1. Keyboard Signal Generation Board

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 487–489, 2010.

The working process of FCR interface has five steps. At first, it detects the face region by Haarcascade method from the captured frames. At second, it detects face components, eyes and mouth. At third, it sets up the face components as the region of interest (ROI). At fourth, it tracks the ROI and recognizes the input states, the smile, the Ho, the wink with right eye, and the wink with left eye in Figure 2. At fifth, PC sends the recognized input state to the keyboard signal generation board by RS-232C. Finally, the board enters the keyboard signal according to the received command into PC.

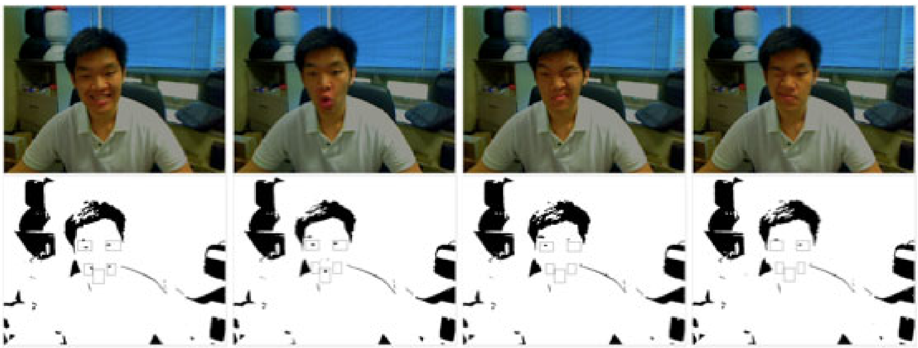


Fig. 2. Face Components as game input (a) Smile, (b) Ho, (c) Wink with left eye, and (d) Wink with right eye

3 Game Description

We have suggested one offline game, *Smile like a Hollywood Star*. It uses face components as game input, directly. Figure 3-(a) shows the screen shot of *Smile like Hollywood Star*. The game is very simple and intuitive. Hollywood star's photo is displayed and a user mimics the face expression of the displayed photo like the Hollywood star. When the feature points, face components, are similar between the player and the star, the player gets the game score. The game score is proportional to the completion time.

Fig. 3. Games with Face Components Recognition (FCR) Interface (a) *Smile like Hollywood star* (b) Kart-Rider serviced by NEXON, and (c) Tetris by NHN

Besides offline game, we have applied FCR interface to two commercialized online games in Korea. One is Tetris [4] in Figure 3-(b) and the other is Kart-Rider [5] in Figure 3-(c). In Tetris, the smile rotates a block, the Ho fast drops a block, the wink with right eye moves a block to the right, and the wink with left eye moves a block to the left. In Kart-Rider, the smile drives the Kart to forward, the Ho drives the Kart to backward, the wink with right eye turns right the Kart, and the wink with left eye turns left the Kart.

4 Conclusion

In this study, we suggest new game interface that uses face components as game input, directly. By adopting the suggested FCR interface, we have presented one new types of game, *Smile like Hollywood Star*, and FCR interface has applied to two online games.

FCR interface and three games have three purposes. One is that the users have more natural smiling face. Usually, the East Asian is impassive than the Westerner, because they feel awkward to express an emotion [6]. By iteratively playing the game with FCR interface, we expect that the user feels more comfortable to reveal an emotion and have dynamic facial expression including a smiling face. Another purpose is reducing the level of stress and promoting a positive emotion by smile, even though it is artificially smile. The other is the game input style for mobility impaired people and hand impaired people.

References

1. Bernhaupt, R., Boldt, A., Mirlacher, T., Wilfinger, D., Tscheligi, M.: Using emotion in games: emotional flowers. In: Proceedings of the International Conference on Advances in Computer Entertainment Technology, pp. 41–48. ACM, New York (2007)
2. Lankes, M., Riegler, S., Weiss, A., Mirlacher, T., Pirker, M., Tscheligi, M.: Facial expressions as game input with different emotional feedback conditions. In: Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology, pp. 253–256. ACM, New York (2008)
3. Obaid, M., Han, C., Billinghurst, M.: Feed the Fish: an affect-aware game. In: Proceedings of the 5th Australasian Conference on Interactive Entertainment. ACM, New York (2008)
4. Tetris, H.: `http://tetris.hangame.com`
5. KartRider, N.: `http://kart.nexon.com/kart/page/Main.aspx?URL=Home/index`
6. Kayan, S., Fussell, S.R., Setlock, L.D.: Cultural differences in the use of instant messaging in Asia and North America. In: Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work, pp. 525–528. ACM, New York (2006)

Study on an Information Processing Model with Psychological Memory for a Kansei Robot

Yuta Kita[1], Masataka Tokumaru[2], and Noriaki Muranaka[2]

[1] Graduate School of Kansai University, 3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan
k366457@kansai-u.ac.jp
[2] Kansai University, 3-3-35 Yamate-cho, Suita-shi, Suita, Osaka 564-8680, Japan

Abstract. In this paper, we propose an information processing model for a kansei robot. This model handles memory based on human psychology. We expect that on incorporating the model, a robot can exhibit human characteristics because of using psychological memory. To verify the model, we first perform a comparison between the results of the experiment performed using this model and that of an actual psychological experiment. The results of the comparison suggest that the memory functions of the model are similar to the human memory functions. Second, we conduct the process of learning movement actions to verify that the robot on which the model was implemented learned movement for moving to many places and decreasing its curiosity.

Keywords: kansei, robot, memory, neural network, psychology.

1 Introduction

For making partner robots resemble humans they need to have human characteristics. However, the method to instill such characteristics in robots is not known yet. Therefore, we suggest an information processing model based on human psychology [1]. This model is constructed by referring to information processing models [2]. The model explains how internal or external information is handled. In this study, we verify the usability of the model, particularly for memory.

2 Information Processing Model

Figure.1 shows the information processing model we propose. First, data captured by the sensors of a robot enters the Sensory Memory and then, the data is transferred to the Short Memory only if the robot observes (*Observe*, a robot action). Second, the robot's action is generated using a multilayer neural network (NN), whose inputs are the input data in the Short Memory and the robot's internal conditions (e.g., desire).

In addition, the indexes of forgetting and remembering memory data are given to the Short Memory and the Episodic Memory. We apply the function proposed by Wayne A. Wickelgren [3] to the index of the Episodic Memory.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 490–492, 2010.

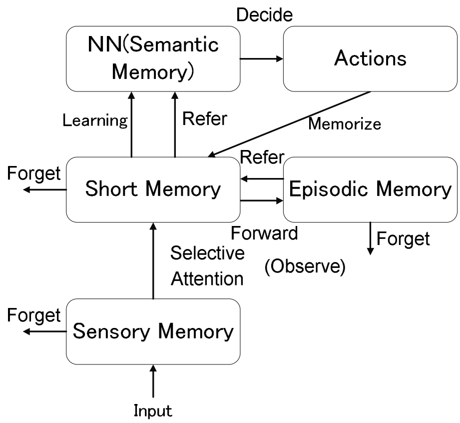


Fig. 1. Information processing model

Here, we define the *want values* which are the values of the robot's internal conditions. The want values steadily increase; they decrease only if a robot's action succeeds. When the robot learns something, the NN is given the spread coefficient referring to the rate of the decrease in the want value. However, this learning is executed only when a robot's action ends, and there is a possibility that this learning is inadequate. Therefore, a robot executes the offline learning when the input data in the Short Memory is similar to data in the Episodic Memory (Recall under Cued).

We performed a simulation for verifying that the memory function of the model is similar to the human memory function. This simulation is a comparison between the outcome of the model and that of the experiment conducted by Bennet B. Murdock [4]. The results indicate that their similarity is 90.6%; thus we assumed that their functions are similar.

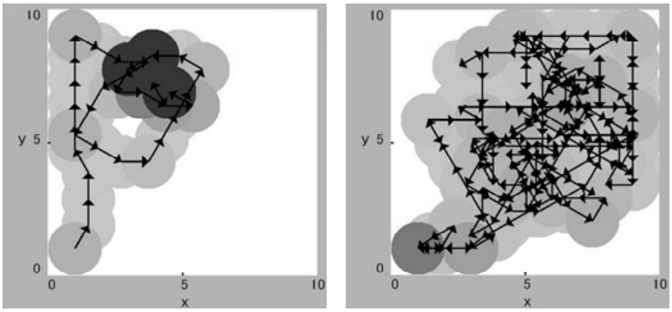
3 Learning the Movement Actions Simulation

From the simulation, we can determine the processes for learning the movement actions. Here, there are six robot actions, generated by the recurrent NN; *turn left* and *right*, *move forward* and *backward*, *observe*, and do *nothing*. The robot's want value is referred to as *curiosity*. Equation (1) shows how to decrease its curiosity if a robot moves.

$$Curiosity_k = Curiosity_{k-1} \times (0.5 + 0.5 \times Sim(position)). \quad (1)$$

Here, *Sim(position)* is the function, which returns the similarity between position information of memory and that of the robot's position; Lesser the memory a robot has regarding the place, the more its curiosity decrease.

Figure.2 shows the path (indicated by arrows) of a round robot moving in a square room with action steps. The figure also shows that the darker the shade of the path is, the more frequent the robot is present at that location. We assumed that the robot



(a) 0-500 steps (b) 1000-1500 steps

Fig. 2. State of the moving round robot

learned the movement actions needed to travel throughout the room to decrease its curiosity, because the robot had moved a considerable number of steps.

4 Conclusions

In this paper, we proposed an information processing model for a kansei robot. In addition, we performed a comparison of the results of this study with the experiment that Bennet B. Murdock conducted, which indicated that the memory function in the memory and the human memory function have a similarity is 90.6%. The second simulation for learning the movement actions indicated that the robot moved to decrease its curiosity.

This study showed that the robot exhibited human characteristics. However, we still need to verify the features of an actual robot because an experiment with an actual robot was not conducted.

References

1. Inoue, K., Tokumaru, M., Muranaka, N.: Considering the Characteristics of Living Creatures for Kansei Robot Action Generation Using a Neural Network. In: Joint 4th International Conference on Soft Computing and Intelligent Systems and 9th International Symposium on Advanced Intelligent Systems, pp. 50–55 (2008)
2. Baddeley, A.: Working Memory, Thought, and Action. Oxford University Press, USA (2007)
3. Wickelgren, W.: Single-trace fragility theory of memory dynamics. Memory & Cognition 2(4), 775–780 (1974)
4. Murdock Jr., B.: The serial position effect of free recall. Journal of Experimental Psychology 64, 482–488 (1962)

Vegetation Interaction Game: Digital SUGOROKU of Vegetation Succession for Children

Akiko Deguchi[1], Shigenori Inagaki[2], Fusako Kusunoki[3], Etsuji Yamaguchi[2], Yoshiaki Takeda[2], and Masanori Sugimoto[4]

[1] Faculty of Education, Utsunomiya University, Japan
deguchia@cc.utsunomiya-u.ac.jp
[2] Graduate School of Human Development and Environment, Kobe University, Japan
inagakis@kobe-u.ac.jp, etuji@opal.kobe-u.ac.jp, takedayo@kobe-u.ac.jp
[3] Faculty of Art and Design, Tama Art University
kusunoki@tamabi.ac.jp
[4] Graduate School of Engineering, University of Tokyo
sugi@itl.t.u-tokyo.ac.jp

Abstract. In this study, we redesign and develop a new digital sugoroku game based on the phenomenon of vegetation succession. A practical evaluation in an elementary school that consisted of game play and fieldwork activity was conducted. The results of the evaluation showed that the game was effective in stimulating the interest of the students who participated in the game, and was able to support their learning in a joyful way.

Keywords: Digital interaction game, vegetation succession.

1 Introduction

Simulation games play an important role in the teaching and learning of science at school. They can support enjoyable interactive learning of complex systems in nature that are difficult to deal with in the school laboratory [1].

In our previous study, we have developed a simulation game, the “Vegetation Interaction Game” based on the vegetation succession, which is a digital sugoroku board game that simulates the real forest area of Mt. Rokko in Japan. In this sugoroku board game, a piece corresponds to a plant, and a grid of sugoroku to the relative succession phase around the plants. In playing the game, each child handles one piece. This condition supports children's playing the role of agent in the simulation. Moreover, we conducted an experimental evaluation in a university in Japan. The results showed the game was effective in stimulating the interest of students, and in supporting their learning in a joyful way[2].

In this study, we made some improvements in game interface to strengthen the enjoyability and understandability of the game. Additionally, we conducted a practical evaluation that consisted of two phases; game and fieldwork activity phase. This issue shows the development and improvement of our “Vegetation Interaction Game,” and the results of practical evaluation.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 493–495, 2010.

2 Development and Improvement of the Game

Figure 1 is the main window of the digital game; it shows the face of the sugoroku board. 6 plant pieces represent 6 characteristic plants that grow in the Mt. Rokko region. The size of the sugoroku board was set at 1024*768 pixels. The surrounding part is the grid area of the sugoroku board. The central part houses the event cards area, a direction window to move pieces, and a visualization window to show vegetation succession according to the progress of the game.

Fig. 1. Main window of sugoroku board

6 players can participate in one game. Each player handles one piece. Players draw event cards, one at a time, by clicking in turn. When a plant piece advances ahead on the board grids, it implies that the plant is dominant in that particular environment. Each piece is to be moved by the number of grids that the current event card indicates. These directions are shown in the central part of the main window. If more than one piece takes the same position on the grid, they will be moved along the grids as indicated by the rules governing the interaction between plants. A game finishes when all event cards are drawn.

In this study, following points were improved based on the feedback of the previous experimental evaluation; (a) layout of the interface to make it more understandable, (b) illustration of event cards and visualization window to make them more scientifically valid and enjoyable for children.

3 Practical Evaluation of the Vegetation Interaction Game

Fig. 2. Game Playing

Fig. 3. Field work activity

The subjects were 17 sixth grade children (12 years old) in a Japanese elementary school. We recruited 18 children to participate in the experiment, and the participants were divided into three groups (6, 6, and 5) in random order (one child was absent on the experiment day). Game play was held in the elementary school (Fig. 2). Children played the game three times; in each game, the numbers of each event card were controlled so that children could experience some patterns of vegetation succession. After playing the games, the children had a fieldwork activity in Mt. Rokko area experiencing real vegetation succession along with the guidance of phytosociologist who is proficient at Mt. Rokko area (Fig. 3).

3.1 Evaluation of Students' Interests

After the game play, children answered the questionnaire about the game. Table 1 shows the question items and their results. To examine the children's answers, "think so" and "quite think so" were evaluated as a positive answer, "no opinion," "don't quite think so" and "don't think so" were evaluated as a negative answer. After that, Fisher's exact test was conducted. The results show that positive answers were significantly higher than negative ones in all the items.

Table 1. Results of students' interests

Questions	TS	QTS	NO	DQTS	DTS
1) I enjoyed playing the game.**	13	4	0	0	0
2) I was absorbed in playing the game.**	15	2	0	0	0
3) I understand the relation between disturbances and the number of the kinds of plants.**	10	5	0	2	0
4) I understand the characteristics of interactions in vegetation succession.**	10	7	0	0	0

TS; Think so, QTS; Quite think so, NO; No opinion, DQTS; Don't quite think so, DTS; Don't think so, ** $p<0.1$

4 Conclusion

In this study, we redesigned and developed a Vegetation Interaction Game and conducted a practical evaluation that include fieldwork activity. The results of evaluation showed that children who participated in the experiment felt the game was interesting and they also thought that they understood the mechanisms of vegetation succession. It was shown that the game could successfully communicate with students, and was able to support their understanding of the relationship between disturbance and the numbers of each plant.

Acknowledgement

This research is supported by Grants-in-Aid for Scientific Research (B) for Masanori Sugimoto, and Hayao Nakayama Foundation for Science & Technology and Culture for Akiko Deguchi.

References

1. Squire, K., Klopfer, E.: Augmented reality simulations on handheld computers. The Journal of the Learning Sciences 16(3), 371–413 (2007)
2. Deguchi, A., Inagaki, S., Kusunoki, F., Yamaguchi, E., Takeda, Y., Sugimoto, M.: Development and Evaluation of a Digital Vegetation Interaction Game for Children. In: Natkin, S., Dupire, J. (eds.) Entertainment Computing – ICEC 2009. LNCS, vol. 5709, pp. 288–289. Springer, Heidelberg (2009)

Penmanship Learning Support System: Feature Extraction for Online Handwritten Characters

Tatsuya Yamaguchi[1], Noriaki Muranaka[2], and Masataka Tokumaru[2]

[1] Kansai University Graduate School, 3-3-35 Yamate-cho, Suita-shi, Osaka, 564-8680 Japan
[2] Kansai University, 3-3-35 Yamate-cho, Suita-shi, Osaka, 564-8680 Japan
{k526329,muranaka,toku}@kansai-u.ac.jp

Abstract. This paper proposes a feature extraction method for online handwritten characters for a penmanship learning support system. This system has a database of model characters. It evaluates the characters a learner writes by comparing them with the model characters. However, if we prepare feature information for every character, information must be input every time a model character is added. Therefore, we propose a method of automatically extracting features from handwritten characters. In this paper, we examine whether it correctly identifies the turns in strokes as features. The resulting extraction rate is 80% and in the remaining 20% of cases, it extracted an area near a turn.

Keywords: Penmanship, Evaluation of characters, Features, Turns.

1 Introduction

Classes and correspondence courses on penmanship are very popular. In the penmanship classes, the learner attends there regularly. Therefore, the time to attend there is needed, and it is not suitable for a busy person. In a correspondence course, the learner learns the shape and balance of a character according to a model character in the text and the accompanying explanation at favorite time. Afterward, the learner sends the teacher his fair copy of the character, and the teacher corrects it and sends the result back to the learner. This is the general flow of a correspondence course, which lacks real-time feedback due to the time required for sending materials. Therefore, we developed a penmanship learning support system that automatically evaluates characters. The purpose of this system is to reproduce the environment of the penmanship classroom at home. In other words, the learner can learn penmanship freely without being limited in place and time. In this research, we extract features from the model character and that produced by the learner. We then evaluate the characters by comparing the length of features and the angles between them. In this paper, we describe the feature extraction method. We want to extract the beginnings, endings, turns, and curves of the strokes. We do not describe the hardware and software here because we describe them in [1].

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 496–498, 2010.

2 Feature Extraction

The character data has coordinates at 10 ms intervals. We call the each of the points located at these intervals a Point. We extract the points that become feature candidates from the Points and then define the selected Points as features.

First, we describe the method of extracting the points that are feature candidates. It can be divided into three stages. In the first stage, we let the beginning and ending points P_B and P_E, respectively, of the stroke be feature candidates. In the second stage, we search for the feature parts. We describe the procedure below.

1-1. Replace P_B with a base point P_b.

1-2. Replace the second Point from P_b with a moving point P_m.

1-3. For all Points that exist between P_b and P_m, evaluate the distance D_P between it and the straight line P_b P_m. Evaluate the Point $P_{\max}$ and the distance $D_{\max}$ for which D_P is the maximum.

1-4. If $D_{\max}$ is more than δ_1 (which is 2 in this paper), let $P_{\max}$ be a feature candidate point. Replace $P_{\max}$ with P_b. If the second Point from P_b is P_E, end. Otherwise, move to step 1-2.

1-5. Otherwise, P_m is moved to the following Point. If P_m is P_E, end. Otherwise, move to step 1-3.

In the third stage, we let the turns in a stroke be feature candidate points. We show the extraction method below.

2-1. Evaluate $\angle FC_i P_{ij} FC_{i+1}$ of the adjoining feature candidate points FC_i, FC_{i+1} and Point P_{ij} that exists between them. ($1 \le i \le N_{FC} - 1$, $1 \le j \le N_P$; N_{FC} is the number of feature candidate points, and N_P is the number of Points that exist between FC_i and FC_{i+1}.)

2-2. For all *j*, find P_{ij} satisfying $\angle FC_i P_{ij} FC_{i+1} < \delta_2$. ($\delta_2$ is a threshold; a value of 100 is used in this paper.) If it is not found, move to step 2-5.

2-3. Let distance $D_{P_{ij}}$ denote the minimum of the two distances between FC_i and P_{ij} and between FC_{i+1} and P_{ij}.

2-4. Let P_{ij} for which $D_{P_{ij}}$ is the maximum be feature candidate points.

2-5. For all *i*, repeat steps 2-1 to 2-4.

Next, we describe the method of deleting an unnecessary feature candidate point. We examine three consecutive feature candidate points FC_{i-1}, FC_i, and FC_{i+1} and delete FC_i that matches the following deletion condition:

3-1. $\angle FC_{i-1} FC_i FC_{i+1}$ is more than 100°.

3-2. The distance between FC_i and the straight line FC_{i-1} FC_{i+1} is less than seven pixels.

3 Experimental Results

We extracted features from the 46 hiragana characters. Fig. 1 shows an example of the resulting feature extraction from three characters. Table 1 shows the resulting extraction of turns in strokes.

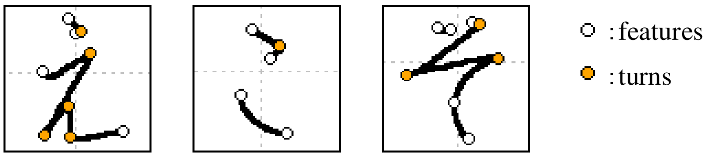


Fig. 1. Example of feature extraction result (“え,” “こ,” and “そ” from the left)

Table 1. Feature extraction of turns in strokes

Number of turns in strokes	Number of extractions	Extraction rate (%)	Average distance of the error extraction (pixel)
80	64	80	1.88

The average distance of the error extraction is the average of the difference in the distance between the turns of 16 features extracted by mistake and the actual features. It can be said that the turns will almost be extracted because they are gaps of less than two pixels.

4 Summary

In this paper, we proposed a feature extraction method for online handwritten characters and examined the turns in strokes that it extracted. The resulting extraction rate is 80%. In the remaining 20% of the cases, it has extracted the area near a turn in a stroke (the average error distance is 1.88 pixels).

Reference

1. Sawamoto, T., Hayano, Y., Muranaka, N., Tokumaru, M.: The Penmanship (script learning) Support System. In: Human Interface Symposium 2008, pp. 383–388 (2008) (in Japanese)

Development of Wall Amusements Utilizing Gesture Input

Takahisa Iwamoto, Atsushi Karino, Masayuki Hida,
Atsushi Nishizaki, and Tomoyuki Takami

Faculty of Information Science and Arts, Osaka Electro-Communication University,
1130-70 Kiyotaki, Shijonawate, Osaka, 575-0063 Japan
{w06018ti,w06039ak,w06094mh,w05098an}@dg.osakac.ac.jp,
takami@isc.osakac.ac.jp

Abstract. Focusing on an infrared camera and a near-infrared radar, we have developed a system for new amusements which we can operate by gesture input toward a screen or moving our bodies in front of a screen projected on a wall. The infrared camera is used for operations performed by means of gestures, and the near-infrared radar is used for operations performed by larger movement of a human body by detecting the position of the person or the state of his feet near the floor surface. A screen is projected on a wall by an ultra-short throw projector. This system can be set up anywhere if there is a certain sized wall.

Keywords: Multi-touch, Digital game, Infrared camera, Infrared radar.

1 Introduction

Studies of gesture input have been proceeding as a method of intuitive operations. Examples for the studies include touch panels, graphics tablets, Apple's Magic Mouse for computers and game operations by Wii remote control toward a TV screen. However, these examples of gesture input are basically used within a restricted range, and unsuitable for large-scale amusements such as physical amusements. Consequently, focusing on an infrared camera and a near-infrared radar, we have developed a system for new amusements which we can operate by gesture input toward a screen or moving our bodies in front of a screen projected on a wall.

2 Devices Used for Gesture Input

An infrared camera and a near-infrared radar are used as devices for gesture input. They have advantages that they are easy to set up because of their compact sizes and they have sufficient processing capacities. The infrared camera is used for operations performed by means of gestures, and the near-infrared radar is used for operations performed by larger movement of a human body by detecting the position of the person or the state of his feet near the floor surface.

An infrared camera, OptiTrack FLEX V100 is used. The data of three-dimensional coordinates, sizes and areas of markers are obtained by irradiating infrared light from

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 499–501, 2010.

around a lens on the camera and taking the reflected infrared light from the surface of the markers. An image taken by the camera has only the markers and the other part of the image except the markers such as background and others becomes "transparent". Processing of the captured image is immediately performed and gesture input is immediately reflected in output image of an application. Image capture is performed 100 times per second, a view angle is 45-degree and the range of capture is 3 cm to 6 m when a 1-inch standard marker is used (figure 1).

The near-infrared radar rotates laser beams 36 times per second with a 0.36-degree pitch in fan-shaped flat space with radius 4 m and central angle 240-degree, and can acquire the data of directions and the distances to objects by detecting the reflected laser beams from the surface of the objects. The positions of the objects are calculated from these data (figure 1).

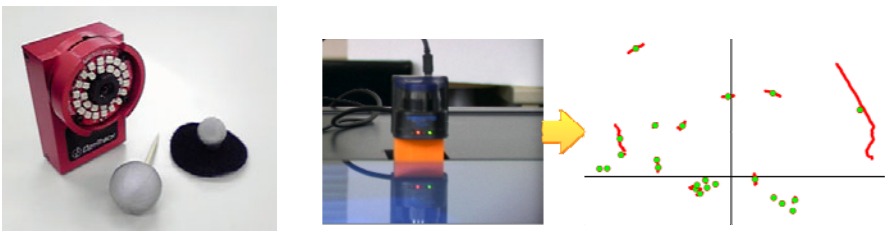


Fig. 1. An infrared camera and markers(left) and data detected by a near-infrared radar(right).

3 System Structure

A screen is projected on a wall by an ultra-short throw projector. This system can be set up anywhere if there is a certain sized wall. Further, as it is assumed that an operator stands in front of a screen, there is a possibility that a body silhouette appears on the screen when normal projectors are used, however, if an ultra-short throw projector is used, the silhouette does not almost appear on the screen because the main body of the projector can be placed almost adhered to a wall.

The infrared camera is fixed with a tripod and installed in the position at the same height or a little higher than the operating position. The camera is connected to a personal computer with an USB cable and captured images are sent to a server using API. The server performs image processing, calculates the coordinates, height, width and area of each detected marker, lumps the data together and sends it to an application, and the data is utilized for the application. The near-infrared radar is set up on a floor and detects the standing position and other data of an operator. Figure 2 shows the system configuration.

The recognition of the position of a person standing on the floor is performed by placing the near-infrared radar parallel to the floor. The radar can recognize the shapes of the person as follows; if both feet are detected by the radar, the person is standing, if one foot is detected by the radar, the person is standing on one foot and if the radar does not detect anything, the person is jumping. Squat is recognized by placing two radars and sensing at different heights. This system can be available for several people at the same time.

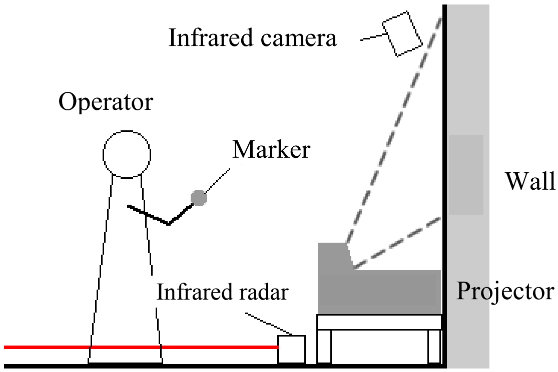


Fig. 2. System configuration

4 An Example of Applications

A product example of applications using gesture input by an infrared camera is shown in Figure 3. This is an amusement in which cards displayed on a wall are moved by gestures. The card overlapped with a marker goes away at random and an image hidden behind these cards appear.

Fig. 3. A product example of an application using gesture input by an infrared camera

Study on an Emotion Generation Model for a Robot Using a Chaotic Neural Network

Hiroyuki Sumitomo[1], Masataka Tokumaru[2], and Noriaki Muranaka[2]

[1] Graduate School of Kansai University, 3-3-35 Yamate-cho Suita-shi, Osaka, 564-8680 Japan
[2] Kansai University, 3-3-35 Yamate-cho Suita-shi, Osaka, 564-8680 Japan
k067993@kansai-u.ac.jp

Abstract. This paper proposes an emotion-generation model for complex change using a chaotic neural network (CNN). Using a CNN, the proposed model will solve the problem of past studies that have indicated that robotic emotion changes are simplistic. The model uses the principle of an adaptation level, which is used in Russell's emotion model to generate emotion. This paper considers the effectiveness of this approach using simulation, and shows that the model can express a change of "adaptation". In addition, through the chaos of CNN, the proposed model can express different changes, even if the values of CNN's input values remain the same.

Keywords: Partner Robot, Chaos Neural Network, Emotion.

1 Introduction

Recently, the market for partner robot has expanded with along with technological advances. A partner robot is a robot that receives communications from a person and brings healing to the person's mind. However, because the expressions of emotion that it brings are simple, both interest and research studies on this subject have waned. This study examines, using a chaotic neural network (CNN), a model that can generate complex emotions that are similar to the emotions of human beings.

2 The Emotion-Generation Model

The emotion-dimension theory is used to construct the emotions of a partner robot. In this paper, Russell's emotion model can express emotion by two dimensions consisting of two axes, the pleasant level and the awake level [1]. Russell's emotion model also exhibits the principle of adaptation. It shows that an emotion is effected by the emotion of the previous event. For example, a glad event when a emotion is a little sad is felt gladder than usually. In this paper, CNN is used to show a complex output change [2]. This is followed by the details of the emotion-generation model, and the network structure of CNN, which uses a layered network.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 502–504, 2010.

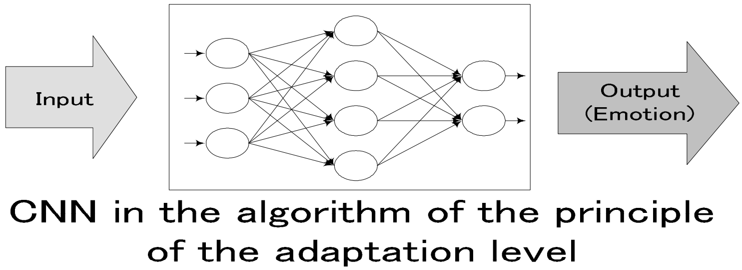


Fig. 1. Structure of the proposed model

$$y(t) = \eta(t) + \zeta(t) \tag{1}$$

$$x(t) - 0.42x(t-1) = \frac{2}{1+\exp(-(\eta(t)+\zeta(t)+\beta(t))/\varepsilon)} - 1 \tag{2}$$

$$s = 1 + \exp(-(\eta(t)+\zeta(t))/\varepsilon) \tag{3}$$

$$y(t) = -\varepsilon \log\left\{\frac{s}{1-0.21sx(t-1)} - 1\right\} \tag{4}$$

In addition, the study method uses a back-propagation algorithm. A structural drawing of the proposed model is shown in Figure 1. To express complex affective changes, the proposed model includes an algorithm that exhibits the principle of the CNN adaptation level. The expression in an internal state of the neuron of CNN is (1). For the proposed model, adding the principle of adaptation level to (1), (1) changes (3) and (4), using (2).

The proposed model prepares five inputs. Two of them take either −1 or 1 at a constant probability, and neither values becomes 1. Another two inputs change the probability that the value is −1 or 1, according to other inputs. These four values do not change for a period that is decided at random. The remaining input takes the value of two significant figures for −1 to 1. In addition all inputs take a gradual time-series change.

3 Simulation

In this simulation, the proposed model shows the feature of the adaptation level, and we confirm whether the changes in output of the proposed model are more complex than that of CNN. At first, both models study enough with a same input value.

Next, the output changes of both models were compared when they have the same input value. The results of five trials are shown in Figure2 and 3.

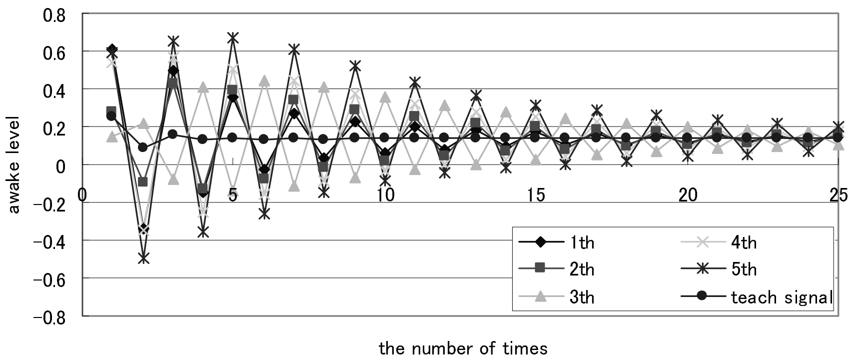


Fig. 2. Output change of the proposed model

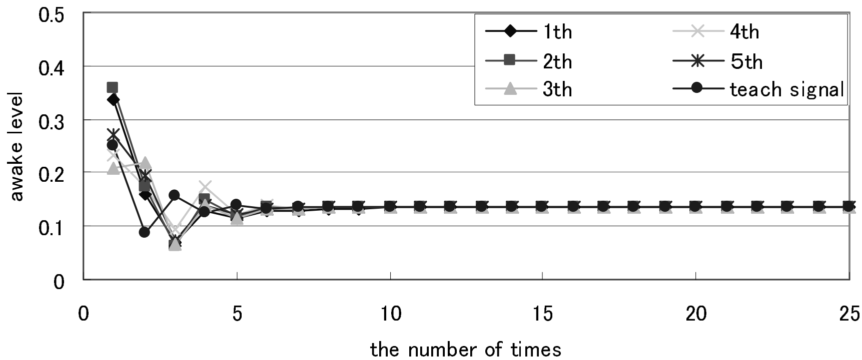


Fig. 3. Output change of CNN

Figure2 and 3 show an awake level's changes in Russell's emotion-generation model.

For the proposed model, the output change increases and decreases alternately as the value receives the influence the principle of the adaptation level. However, in CNN, it was almost steady with the value of the teach signal.

Moreover, the proposed model can clearly confirm how the output can be settled differently at each trial, although the chaos generated because of the initial weight values of neurons are different. Therefore, we can confirm that the proposed model shows more complex changes of emotion than CNN.

I will evaluate an emotion of the robot with the proposed model.

References

1. Hama, H., Suzuki, N., Hama, Y.: Invitation to emotion psychology. In: Approach to feelings and emotion, Saiensu-Sya Ltd.
2. Gouhara, K.: Chaos of neural system. Tokyo Denki University Publication

Interactive Tabu Search vs. Interactive Genetic Algorithm

Tatsuya Hirokata[1], Masataka Tokumaru[2], and Noriaki Muranaka[2]

[1] Gradute School of Kansai University.3-3-35, Yamatre-cho, Suita-shi, Osaka, Japan
[2] Kansai University.3-3-35, Yamatre-cho, Suita-shi, Osaka, Japan
{k875239,toku,muranaka}@kansai-u.ac.jp

Abstract. We propose an interactive tabu search (ITS) to be used for the development support of a product that fits a human's feeling. Interactive evolutionary computation (IEC) is one of the technologies used in the development support of products that fit a human's feeling using a computer and person undergoing a communication. The interactive generic algorithm (IGA) is generally used in the IEC. A major problem with the use of the IEC is the increased burden on the IEC user to evaluate multiple solution candidates. Using the ITS instead of the IGA may reduce this burden, because the ITS user chooses only his most favorite solution candidate among multiple solution candidates. We performed a comparison of the search performance using simulations with the ITS and IGA. As a result of this simulation, the search performance of the ITS exceeded that of the IGA by a range from 2% to 10%.

Keywords: Tabu search, Interactive evolutionary computation.

1 Introduction

Interactive evolutionary computation (IEC) is a technique used in the development support of a product that fits a human's feeling using a computer.

A major problem with the use of the IEC is the increased evaluation burden of the IEC user. One of the methods that can be used to reduce the evaluation burden of the IEC user is an improvement to the evaluation interface [1]. A general evaluation interface of IEC is an *n*-stage evaluation method of all of the individuals that the IEC system shows. However, if many individuals are shown immediately, the result is a large evaluation burden for the IEC user.

The burden of the IEC user is caused as follows. First, the user should evaluate the same individual repeatedly during a single search. Second, the user must give each individual a gradual evaluation point. To solve these problems, an algorithm that can maintain evolution performance by a simple evaluation interface is desirable. However, this is difficult to achieve when using a generic algorithm (GA).

Therefore, we use a tabu search (TS) instead of a GA. The ITS aims to reduce the evaluation burden of the IEC user.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 505–507, 2010.

2 Interactive Tabu Search

The ITS is a tabu search that evaluates the solution candidates using human intuitive evaluation instead of a quantitative evaluation function. The evaluation interface of the ITS is a method to choose the individual that the user prefers the most. We anticipate that the simplicity of this solution evaluation greatly reduces the burden of the user.

Following are the steps of the ITS process:

(1) Generation of the initial individuals.
The initial individuals are generated by the bit strings at random.
(2) Presentation and evaluation by the users.
All of the individuals are shown to the user, who chooses the individual that he prefers the most.
(3) Update of the tabu list.
The information of the individual that the user has chosen is added to the tabu list.
(4) Generation of neighboring individuals.
The neighboring individuals are generated by flipping the single bit individual that the user chose. Multiple neighboring individuals are generated. Steps (2) to (4) are repeated until the final generation.
(5) Generation of the initial individuals from the tabu list.

The TS searches while repeating improvement and the change for the worse of the solution. Therefore, the evaluation of an individual chosen before several generations may be higher than the evaluation of the current individual. The tabu list records the chosen individuals from several generations. The individuals in the tabu list are shown to the user.

3 Simulation and Result

We evaluate the utility of the ITS with the use of a simulation, which compares the search performance of the ITS with the IGA. The utility of the ITS can be proven if the search performance of the ITS is better than that of the IGA.

The target individual to mean human's feeling is generated, and evaluates the individual in Hamming distance with it. Figure 1 shows the evaluation of the individuals according to the target individual.

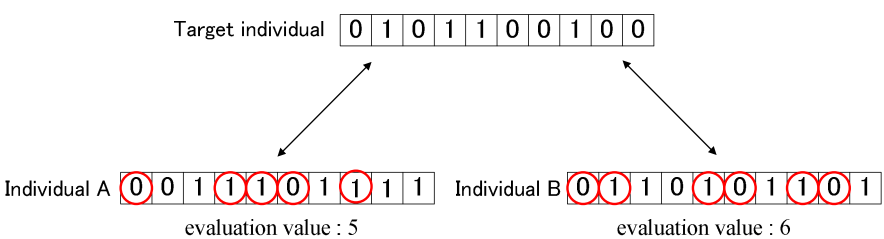


Fig. 1. Evaluation of the individual by the target individual

The evaluation value is the number of bits that a number accords with it. If the length of the bit string is longer than 10 bits, the evaluation value is converted so that the maximum length becomes 10.

Figure 2 shows a result of the simulation. As a result of this simulation, the search performance of the ITS exceeded the performance of the IGA by approximately 2% with the 20th generation of 10 bits. In addition, the search performance of the ITS exceeded that of the IGA by approximately 10% with the 20th generation of 50 bits. The ITS achieves the same evaluation value with half the generation number of the IGA. The ITS can be figured to be algorithm that can maintain search performance by a simple evaluation interface. So we consider that the ITS be expected to reduce the evaluation burden of the IEC user.

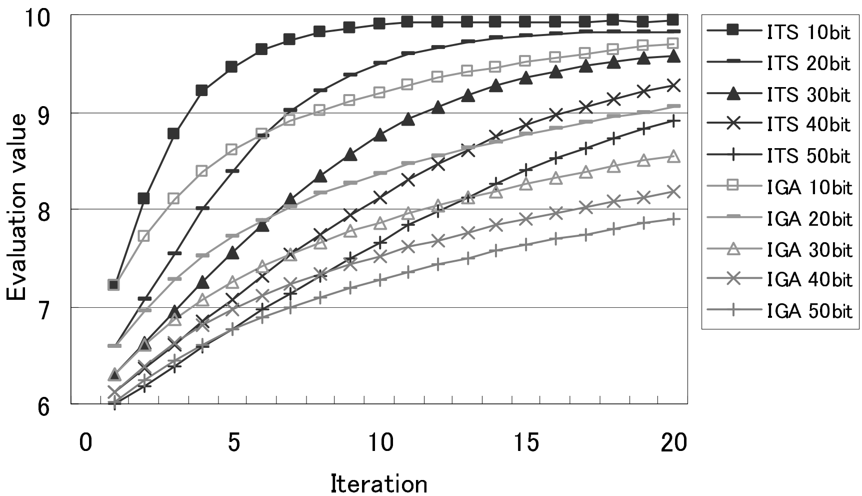


Fig. 2. Search performance of ITS and IGA

4 Summary

We propose the use of an ITS to reduce the burden of users. The utility of the ITS was proved by simulation. In the future, we will inspect the evaluation burden of the user who is used as a subject.

Reference

1. Takagi, H.: Interactive Evolutionary Computation: Fusion of the Capabilities of EC Optimization and Human Evaluation. Proceedings of the IEEE 8(9), 1275–1296 (2001)

Instant First-Person Posture Estimation

Takafumi Serizawa[1] and Yasuyuki Yanagida[2]

[1] Graduate School of Science & Technology, Meijo University
[2] Faculty of Science and Engineering, Meijo University
1-501 Shiogamaguchi, Tempaku-ku, Nagoya, 468-8502 Japan
m0941501@ccalumni.meijo-u.ac.jp,
yanagida@.meijo-u.ac.jp

Abstract. We propose an instant posture estimation technique, which can operate only by a stereo image with small (6 cm) baseline. It does not require a priori information of the target user, background information, and markers. This system allows cameras to move freely, because it operates only with a small stereo camera unit. Moreover, if the input image is replaced with a movie or real-time video, this system can be used as a real-time motion tracker. With our proposed technique, robots and computers will be able to non-verbally communicate with unspecified people as well as pre-registered people. Moreover, this system can output not only the posture but also the body size and clothes. Therefore, proposed technique can also be used as a calibration procedure for other motion tracking algorithms.

Keywords: posture estimation, motion tracking, Marker-less, Calibration-free.

1 Introduction

The motion tracking technology is necessary to achieve a natural interaction of human and computer/robot as will as recognition of voice and facial expression. Current "motion tracking" technologies based on volume matching or image matching have achieved high-speed and high-accuracy. However, existing techniques require a priori information of the person being tracked [1]. When the target is a specific user, a priori information of the target user can be input before the tracking session. But, when the user is unspecified, it is impossible to input a priori information. To solve this problem, it is necessary to develop a new algorithm to estimate initial information based on the input of single frame.

This paper explains the "posture estimation" algorithm to acquire initial information and the entertainment application using our algorithm.

2 Proposed Technique

We propose a robust algorithm of posture estimation (Fig. 1), by avoiding error propagation from body trunk to ends of limbs and by using reliable 3D body ridge

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 508–510, 2010.

line detection in estimating the position/orientation of each part of the body. To avoid the error propagation, we introduced loose coupling of bones and a mechanism to feedback the estimation result of upper arms to the rotation angle of body trunk (Fig. 2). Unlike existing methods that use rigid joint-bone structure and estimate the direction/rotation of each part serially, estimation error in the rotation angle of body trunk does not propagate to the errors at the end of limbs (Fig. 3).

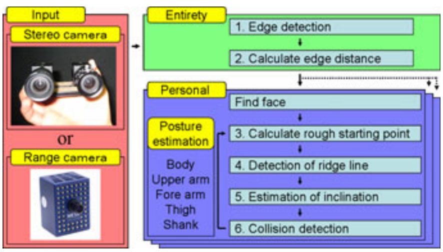


Fig. 1. System

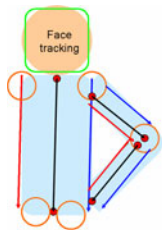


Fig. 2. Proposed

Fig. 3. Related

3 System Performance

Fig. 4-7 shows the case that parts of the body (lower legs) are out of the camera view, with complex background. Fig. 8-11 shows the case that entire body is visible with simple background.

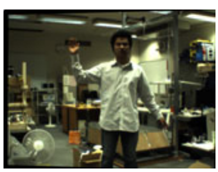


Fig. 4. Left input A

Fig. 5. Right Input A

Fig. 6. 2D Output A

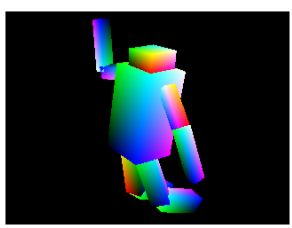


Fig. 7. 3D Output A

Fig. 8. Left input B

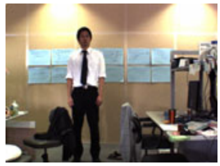


Fig. 9. Right Input B

Fig. 10. 2D Output B

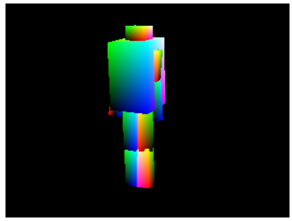


Fig. 11. 3D Output B

4 Our Demonstration

In our demonstrations, a visitor (A) is asked to put on a headgear composed of an HMD and a stereo camera. When another person (B) enters the visitor's view, the appearance of B is replaced by a 3D character (Fig. 12).

Fig. 12. Demonstration (3D model and character (C) Teatime.inc)

Reference

1. Olivier, B., Pascal, C.M.C., Arnaud, B.: Fast nonparametric belief propagation for real-time stereo articulated body tracking. Computer Vision and Image Understanding 113(1), 29–47 (2009)

Monitoring User's Brain Activity for a Virtual Coach

Bram van de Laar, Anton Nijholt, and Job Zwiers

Human Media Interaction, University of Twente, Enschede, The Netherlands
{laar,nijholt,zwiers}@ewi.utwente.nl

Abstract. The system described in this paper is an attempt at developing a coach for sports using a virtual world and multimodal interaction, including brain activity. Users can ride a bicycle through a virtual world while the coach monitors the user's performance. The system incorporates the user's brain activity, heart rate and respiration rate. These data are analyzed and the features are sent through to give the virtual coach the instructions for movements and dialogues to coach the user. The Electroencephalogram (EEG) provides ample possibilities to research the brain activity of the user and to provide for an extra modality in the interaction.

Keywords: BCI, Coaching, 3D Avatar, Multimodal Interaction, Sports, Extreme EEG.

1 Introduction

Finding motivation for the practice of sports and bodily exercises is becoming harder and harder in the fast-paced society we live in. A lot of people would therefore be helped by the aid of a virtual coach [3]. Earlier work by Ruttkay et al. [7] provides the basis for virtual coaches. The system proposed in this paper is a start for such a virtual coach using brain activity as a modality. In the first working version of this system we focus on a bicyclist, exercising on a real bicycle that has been mounted onto a VR bike frame.

In this system we intend to use brain activity, various physiological signals such as heart rate, respiration rate and volume as well as pedal rate and power exerted by the user on the bicycle.

While heart and respiration rate are quite common features to measure on what level the user is performing, brain activity is not. However, research by Palva et al. [6] and Schürmann et al. [8] indicates that interesting meaning can be found in the alpha band of the electroencephalogram (EEG). Energy in the alpha band can indicate how conscious and attentive the subject is to a task. Also the work by Klimesch et al. [5] provides interesting cues for more research into different energy bands in the EEG of an exercising user. In future research related to this paper we will investigate the use of such information for a better informed virtual coach.

2 Overview of the System

The system as a whole is build up from different components as depicted in figure 1. These components constitute to a loop in which sensory data from the user is

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 511–513, 2010.

measured, analysed and behaviour and dialogues for the virtual coach selected. After this the avatar component realizes the behaviour and dialogue which the user can see on a screen and hear through the loudspeakers.

The first step in the loop is to take the EEG and physiological data. These are acquired from various sensors, located on the user's body. The data is then streamed in real-time to the signal analysis component. This component interprets the data and sends the features to the behaviour selection component. Depending on the value of the incoming features this component selects the most fitting behaviour for the coach. Then in turn, the dialogue system selects the dialogue to send to the Virtual coach. The virtual coach carries out the dialogue and movements which are the feedback for the user.

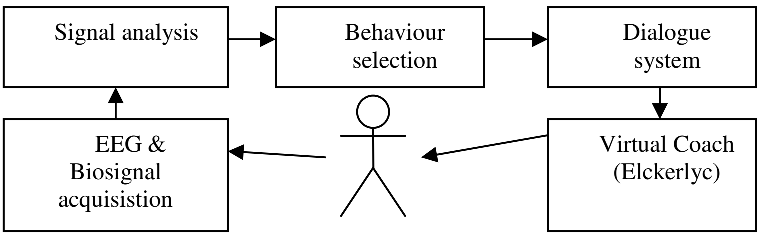


Fig. 1. Overview of the Virtual Coach system

The user can ride through any virtual world which supports the VR bicycle frame. Users sitting on this bike can then use the steer and pedal to navigate through this world, while being monitored and coached by the system. The avatar used for the virtual coach consists of the Elckerlyc system as introduced by Van Welbergen et al. [9]. The Behaviour selection and Dialogue component communicate through Behaviour Markup Language (BML). Elckerlyc visualizes the movements and dialogues on the screen.

3 Analysis and Use of Neurophysiological Data

In the current version of the system the EEG is analysed for energy in the alpha band on a certain part of the brain, namely location ‘Pz’ in the 10-20 placement system [2]. The feature measured on this location on the scalp is an indication for the user either being agitated and alert or relaxed and distracted. The EEG is band pass filtered between 8 and 13 Hz after which the band power is calculated. A moving window of 120 seconds is used to normalize the output, id est, the extrema and the mean of the signal over the last 120 seconds determine the actual value of this feature. This way, the algorithm will adapt to any user using the system, invariable of the baseline power in the alpha band of a user.

The interpretation of the user's mental state is then used to select the appropriate attitude for the virtual coach towards the user. The heart and respiration rate will be used to monitor if the user is performing on the right level, depending on the goal being weight loss or condition improvement.

Because the user is performing physical activity, robustness of the EEG signal quality and classification of said signal is very important. However, similar research by Katsis et al [4] in race car drivers indicates that this should be possible.

4 Conclusion and Future Work

In this paper a system for a virtual coach using brain activity is introduced. Research into how brain activity can be put to use for a virtual coach is one of the main interests.

While currently the system only uses one feature in the EEG, future research will include using more complex or multiple features to provide the virtual coach with a more comprehensive representation of the user's mental and physical state. Such features could focus on measuring the flow of the user as defined by Cziksentmihalyi [1]. Ultimately this could lead to a system that keeps track of the user being in the flow and adapts itself to provide a better user experience.

References

1. Cziksentmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper & Row Publishers, New York (1990)
2. Homan, R.W., Herman, J., Purdy, P.: Cerebral location of international 10-20 system electrode placement. Electroencephalography and Clin. Neurophys. 66(4), 376–382 (1987)
3. IJsselsteijn, W.A., de Kort, Y.A.W., Bonants, R., de Jager, M., Westerink, J.: Virtual Cycling: Effects of immersion and a virtual coach on motivation and presence in a home fitness application. In: Proceedings Virtual Reality Design and Evaluation Workshop 2004, University of Nottingham, UK (2004)
4. Katsis, C., Katertsidis, N., Ganiatsas, G., Fotiadis, D.: Toward Emotion Recognition in Car-Racing Drivers: A Biosignal Processing Approach. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 38(3), 502–512 (2008)
5. Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T., Schwaiger, J.: Induced alpha band power changes in the human EEG and attention. Neuroscience Letters 244(2), 73–76 (1998)
6. Palva, S., Palva, J.M.: New vistas for alpha-frequency band oscillations. Trends in Neurosciences 30(4), 150–158 (2007)
7. Ruttkay, Z., Zwiers, J., Welbergen, H.V., Reidsma, D.: Towards a Reactive Virtual Trainer. In: Gratch, J., Young, M., Aylett, R.S., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 292–303. Springer, Heidelberg (2006)
8. Schürmann, M., Basar, E.: Update article: Alpha oscillations shed new light on relation between EEG and single neurons. Neuroscience Research 33, 79–80 (1999)
9. Van Welbergen, H., Reidsma, D., Zwiers, J.: A Demonstration of Continuous Interaction with Elckerlyc. In: 3rd Workshop on Multimodal Output Generation (2010)

Developing Educational and Entertaining Virtual Humans Using Elckerlyc*

Dennis Reidsma, Herwin van Welbergen, Ronald C. Paul,
Bram van de Laar, and Anton Nijholt

Human Media Interaction, University of Twente, The Netherlands
dennisr@ewi.utwente.nl
http://hmi.ewi.utwente.nl

1 Introduction

Virtual humans (VHs) are used in many educational and entertainment settings: training and serious gaming, interactive information kiosks, tour guides, tutoring, interactive virtual dancers, and much more. Building a complete VH from scratch is a daunting task, and it makes sense to rely on existing platforms. However, when one builds a novel interactive VH application, one needs to be able to adapt and extend the means to control the VH offered by the platform, without reprogramming parts of the platform. This paper describes Elckerlyc, a novel platform for controlling a VH. The focus is on how to easily extend and adapt the system to the needs of a particular application, without programming.

2 The SAIBA Framework

The SAIBA framework [1] provides a good starting point for designing interactive VHs. The framework defines a three-stage process: *communicative intent planning*, multimodal *behavior planning*, resulting in a specification of the form synchronisation of the behavior that the VH should display, and *behavior realization*, to actually perform the behavior using the embodiment of the VH.

When designing a VH application, one should focus on the intent and behavior planning stages: given the goals of the VH and the context in the interaction with a human user, determine the communicative behavior that the VH should display. To ensure that this behavior can be executed transparently by any behavior realization platform, the emerging SAIBA standard Behavior Markup Language (BML) is being developed [1].

BML provides a general, realizer-independent description of multimodal behavior that can be used to (incrementally) control a VH. BML expressions describe the occurrence of certain types of behavior on an abstract level (gaze, speech, point and beat gestures, lexicalized gestures and facial expressions, and other types) as well their synchronisation.

* This research has been supported by the GATE project, funded by the Dutch Organization for Scientific Research (NWO) and the Dutch ICT Regie.

H.S. Yang et al. (Eds.): ICEC 2010, LNCS 6243, pp. 514–517, 2010.

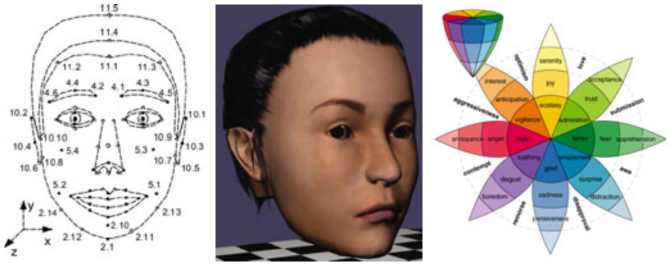


Fig. 1. Face editor: Define, for all MPEG4 control points, the effect on the mesh. Then, obtain subtle face expressions using MPEG4, FACS, or other means.

Although this level of abstraction saves a lot of effort when building VH applications, BML Realizers should never completely be black box systems. One must still be able to specify for some of these abstract behaviors in exact detail how that behavior is performed – for example, by using a specific piece of mocap for a beat gesture, or a hand crafted animation for certain pointing gestures.

3 Elckerlyc

Elckerlyc is a state-of-the-art BML Realizer [2]. Elsewhere, we described its *mixed dynamics* capabilities, that allow one to combine physics simulation with other types of animation in a flexible way [3], and its focus on *continuous interaction*, allowing last moment modification of behavior plans with respect to content and timing, which makes it very suitable for VH applications requiring high responsiveness to the behavior of the user [4]. Here we will discuss how one can adapt Elckerlyc to suit the needs of a particular application without giving up the level of abstraction offered by its BML Realization interface.

Adding New Avatars. Often, VH applications require a custom-built avatar body. Elckerlyc's import facilities allow one to use standard modelling software to build the avatar body. Only two things are needed to control the resulting embodiment using Elckerlyc. Firstly, a *physical model of the avatar* needs to be created using Elckerlyc's physics editor, if one wants to use mixed dynamics. Secondly, in order to have full control of the avatar's face expressions, one needs to *make the face MPEG4 compliant* using Elckerlyc's face editor. For each of the standard MPEG4 control points of a face (see Figure 1), one needs to define their effect on the mesh. Then, all existing face expressions in Elckerlyc can be displayed on the new VH embodiment. Furthermore, the VH can then be controlled using BML Face Behaviors that apply (1) direct MPEG4 control of the face mesh, (2) the FACS Action Units of Ekman et al., which provide a more muscle-like approach, and (3) the basic emotion profiles of [5] using the method of [6].

Extending the Animation Repertoire. Elckerlyc can use motion units (animations) from any number of paradigms, such as motion capture, hand crafted animation, and physical simulation. Each of these paradigms can be seen as providing joint rotations over time, between $0 < t < 1$, with 0 and 1 being the start and end of the animation. For each of the above paradigms, a file format is available from which Elckerlyc can read the specification of new motion units. For facial animation units, the same holds, but instead of joint rotations, a face unit provides displacements of MPEG4 control points over time, between $0 < t < 1$.

Mapping new motion units to specific BML behaviors. After adding new motion units for a specific VH application, the BML Realizer needs to know how to map BML expressions to them. This is achieved through two resource files: the gesture binding and the face binding. Each defines a mapping from BML behavior types to motion units (or face units) in one of the available paradigms, constrained by the value of certain parameters. For example, a <`head`> behavior is mapped to the procedural animation `"nod.xml"` when it satisfies the constraints `action="ROTATION"` and `rotation="NOD"`, and another to motion unit when `rotation="SHAKE"` (see Figure 2).

```
<bml><head id="nod1" repeats="3" action="ROTATION" rotation="NOD"/></bml>
----
<gesturebinding>
 <MotionUnitSpec type="head">
  <constraints>
   <constraint name="action" value="ROTATION"/>
   <constraint name="rotation" value="NOD"/>
  </constraints>
  <parametermap>
   <parameter src="repeats" dst="r"/>
  </parametermap>
  <MotionUnit type="ProcAnimation" file="procanimation/smartbody/nod.xml"/>
 </MotionUnitSpec>

 <MotionUnitSpec type="head">
  <constraints>
   <constraint name="action" value="ROTATION"/>
   <constraint name="rotation" value="SHAKE"/>
  </constraints>
    ...
```

Fig. 2. Gesture binding: link BML behavior types to newly defined animations using the gesture binding, mapping BML parameters to parameters of the motion unit.

4 Conclusion

We globally described how one can use the combination of BML and Elckerlyc to control the behavior of a virtual human, leaving one free to focus on the *communicative intent planning* of the VH rather than the realization of the movements, and we discussed how to adapt and extend Elckerlyc in a very easy way by simply adding new resources.

References

1. Vilhjálmsson, H., et al.: The behavior markup language: Recent developments and challenges. In: Panayiotopoulos, T., Gratch, J., Aylett, R.S., Ballin, D., Olivier, P., Rist, T. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, pp. 99–120. Springer, Heidelberg (2005)
2. Van Welbergen, H., Reidsma, D., Ruttkay, Z.M., Zwiers, J.: Elckerlyc: A BML realizer for continuous, multimodal interaction with a virtual human. To Appear in Journal on Multimodal User Interfaces (2010)
3. Van Welbergen, H., Zwiers, J., Ruttkay, Z.M.: Real-time animation using a mix of dynamics and kinematics. Journal of Graphics Tools (to appear, 2010)
4. Van Welbergen, H., Reidsma, D., Zwiers, J.: A demonstration of continuous interaction with elckerlyc. In: Proc. Multimodal Output Generation (2010)
5. Plutchik, R.: Emotion: A Psychoevolutionary Synthesis. Harper and Row, New York (1980)
6. Raouzaiou, A., Tsapatsoulis, N., Karpouzis, K., Kollias, S.: Parameterized facial expression synthesis based on MPEG-4. Eurasip Journal on Applied Signal Processing 10, 1021–1038 (2002)

Author Index