Stéphane Natkin
Jérôme Dupire (Eds.)

Entertainment Computing – ICEC 2009

LNCS 5709

8th International Conference
Paris, France, September 2009
Proceedings

ifip

Springer

Lecture Notes in Computer Science 5709

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

Stéphane Natkin   Jérôme Dupire (Eds.)

Entertainment Computing – ICEC 2009

8th International Conference
Paris, France, September 3-5, 2009
Proceedings

Springer

Volume Editors

Stéphane Natkin
Jérôme Dupire
Conservatoire National des Arts et Métiers, CEDRIC
292, rue St. Martin, 75141 Paris Cedex 03, France
{stephane.natkin; jerome.dupire}@cnam.fr

Library of Congress Control Number: 2009933197

CR Subject Classification (1998): H.5, H.4, H.3, I.2.1, I.3, J.4, J.5, K.3, K.4

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

ISSN 0302-9743
ISBN-10 3-642-04051-9 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-04051-1 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 SPIN: 12744916 06/3180 5 4 3 2 1 0

*The original version of the book frontmatter was revised:
The copyright line was incorrect. The Erratum
to the book frontmatter is available at
DOI: 10.1007/978-3-642-04052-8_63*

Preface

ICEC 2009 was hosted in Paris, following Carnegie Mellon (2008) and Shanghai (2007). It took place in the CNAM (Conservatoire National des Arts et Métiers), a legendary place of education, research and culture in France and in Europe. The CNAM was founded by l'Abbé Grégoire during the French revolution to deliver the knowledge of science and crafts to everyone ("Omnes Docet Ubique"). It also includes a famous museum of technology, "le Musée des Arts et Métiers," described in the book of Umberto Eco, "Foucault's Pendulum," where one can see the ancestor of all computers "la machine de Pascal" and the first TV set. Today, it is a place where you can find the Computer Research Laboratory, the CEDRIC, where 30 researchers are involved in the field of interactive media and intelligent spaces, the graduate school for games and interactive media (ENJMIN) and where students coming from all around the world can obtain a master degree and a PhD in, for example, game and interactive media or in interaction design. Thus, it is a good place to host the International Conference on Entertainment Computing.

Fig. 1. Machine de Barthélémy (copyright Musée des Arts et Métiers)

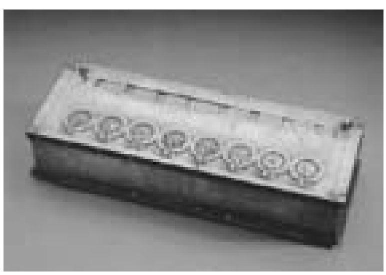


Fig. 2. Machine de Pascal (copyright Musée des Arts et Métiers)

The papers selection process was a hard task: we received 105 submissions coming from 25 countries, most of them having a high quality level. The submission comprised: 39 long papers, 37 short papers, 29 posters and demos. The acceptation rate was, according to each category, 36% (14 long papers), 53% (19 short papers), 77% (23 posters and demos). The themes of these papers cover all the main domains of entertainment computing, from interactive music to games, including a wide range of scientific domains from esthetic to computer science. We were very happy to welcome as keynote speakers six well-known researchers, artists and industrialists from the field of digital art and entertainment: Luc Courchesne, Professor at the University of Montreal and one of the initiators of new media art, Nicolas Gaume, one of the leaders of the French game industry and CEO of Mimesis Republic, Gonzalo Frasca, the famous provocative researcher in the field of game design, Susan Gold, Chairperson of the IGDA Education Sig, Yukiharu Sambe CTO Executive Director at TAITO corporation, the leading Japanese arcade entertainment company, and Michael Stora, the psychologist who uses video games as a therapeutic object. The demo part of the conference, installed in the museum, included a great variety of prototypes and installations demonstrating the creativity of this field. If I try to characterize the evolution of ICEC during the last few years and the 2009 issue, I would point out an increasing number of papers devoted to games and the user's model, a significant number of presentations devoted to sound, and a decreasing number of contributions in the field of computer graphics. The content seems to have become more important than the technology.

I would like to thank all the people and institutions that helped us to realize this conference, in particular: The IFIP as the supporting organization of ICEC and the TC14 Technical Committee; our sponsors, the National Center for Cinema (CNC) and Cap Digital, the Paris cluster in the digital contents field; the universities of Paris 6 and La Rochelle co-organizers of this conference; The AR CNAM Poitou-Charentes in charge of the financial management of the conference; all the people who helped us in the organization of ICEC 2009 and in particular Cecile Le Prado, Daniele Lejais, Jerome Dupire, Ben Salem and all the students of CNAM who were involved in many tasks from designing the web site to helping attendees.

July 2009 Stéphane Natkin

Organization

Conference Chair	Stéphane Natkin	CNAM, France
Program Committee Chairs	Brad Bushman	University of Michigan, USA
	Masayuki Nakajima	Technical University of Tokyo, Japan
Steering Committee	Ryohei Nakatsu	National University of Singapore, Singapore
	Mark Cavazza	University of Teesside, UK
	Zhigeng Pan	Zhejiang University, China
	Stéphane Natkin	CNAM/ENJMIN, France
	Matthias Rauterberg	Technical University of Eindhoven, The Netherlands
	Don Marinelli	CMU, USA
	Hyun Yang	KAIST, Korea
Organization Chair	Jérôme Dupire	CNAM, France
Organization Committee	Françoise Le Vezu	ARCNAM, France
	Danièle Lejais	CNAM, France
	Ben Salem	Technical University of Eindhoven, The Netherlands
Treasurer	Chantal Delaveau	ARCNAM, France
Web Design	Stéphane Gros	CNAM, France
	Xiang Qiu Hou	CNAM, France
	Guillaume Levieux	CNAM, France

Program Committee

Stephen Hansen
Peter Purgathofer
Radslav Yoshinov
Sidney Fels
Zhigeng Pan
David Obdrzalek
Ville-Veikko Mattila
Rainer Malaka
Barnabas Takacs
Richard Reilly
Paolo Ciancarini
Ryohei Nakatsu
Takehiko Kamae
Hitoshi Matsubara
Naoko Tosa
Hyun S. Yang
Matthias Rauterberg
Ben Salem
Jaap van den Herik
Nuno Correia
Adrian David Cheok
Marc Cavazza
Andy Sloane
Lyn Pemberton
Donald Marinelli
Nahum Gershon
Milner Makuni
Tony Manninen
Kevin Wong
Anton Nijholt
Sam Ge
Masayuki Nakajima
Junichi Hoshino
Woontack Woo
Noriko Nagata
Haruhiro Katayose
Kazushi Nishimoto
Hirokazu Katoh
Yoshifumi Kitamura
Tetsuo Ono
Michita Imai
Duh Been-Lirn

Sofia Tsekeridou
Roel Vertegaal
Hyung-Gon Kim
Hisham Bizri
Hiroshi Okuno
Yasuyuki Sumi
James Gimzewski
Victoria Vesna
Inkwon Lee
Gerard J. Kim
Huaqin Shen
Javier Jaen-Martinez
Oscar Garcia-Panyella
Marco A. Gomez-Martin
Tsutomu Terada
Marco Roccetti
Fionnuala Conway
Seiichi Nishihara
Kazunori Miyata
Hiroki Takahashi
Akihiko Shirai
Kaoru Sumi
Masanori Sugimoto
Marcello Carrozzino
Artur Lugmayr
Panos Markopoulos
Elina M.I. Ollila
Lucia Vera
Maria Roussou
Lucia Pannese
Anne Braun
Phillipe Codognet
Pascal Estraillier
Emmanuel Chailloux
Vincent Corruble
Emmanuel Guardiola
Nicolas Szilas
Eric Gressier
Hugues Vinet
Miguel Sicart
Guillaume Hutzler
Jean-Baptiste Labrune
Xavier Reteaux
Valérie Gouet-Brunet
Catherine Pelachaud
Michel Simatic
Pierre Jouvelot
Stefan Gruenvogel
Didier Arques
Craig Lindley
Ludovic Le Bigot
Stephane Donikian
Ruth Aylet
Jean-Claude Lescure
François Pachet
Kristof Berg
Sebastien Deguy
Armelle Prigent
Wang Yansheng
Dominique Archambault
Jesper Juul
Ernest Adams
Jean-Marc Jot
Ian Marshall
Cristina Portalés
Stéphane Gros
Guillaume Levieux
Shuo Hsiu Hsu

Organizers and Sponsors

Conservatoire National des Arts et Métiers
Université de La Rochelle
Université Pierre et Marie Curie

Cap Digital
Centre National de la Cinématographie
IFIP

Table of Contents

Long Papers

Player's Model

Interfaces and Interactions

Sociology of Games

Interactive Storytelling - Interactive Art

Short Papers

Interactive Sound

Design and Experience

Interfaces and Interactions

Interaction Design – Player's model

Programming Interactions

Posters and Demoonstrations

Invited Speakers

Player Performance, Satisfaction, and Video Game Enjoyment*

Christoph Klimmt[1], Christopher Blake[2], Dorothée Hefner[2], Peter Vorderer[3], and Christian Roth[3]

[1] Department of Communication, Johannes Gutenberg University of Mainz, Kleinmann-Weg 2, 55099 Mainz, Germany
[2] Department of Journalism and Communication Research, Hanover University of Music and Drama, EXPO-Plaza 12, 30539 Hannover, Germany
[3] Center for Advanced Media Research Amsterdam (CAMeRA), VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
klimmt@uni-mainz.de

Abstract. An experiment (*N* = 74) was conducted to investigate the impact of game difficulty and player performance on game enjoyment. Participants played a First Person Shooter game with systematically varied levels of difficulty. Satisfaction with performance and game enjoyment were assessed after playing. Results are not fully in line with predictions derived from flow and attribution theory and suggest players to (1) change their view on their own performance with its implications for enjoyment with increasing game experience and (2) to switch strategically between different sources of fun, thus maintaining a (somewhat) positive experience even when performance-based enjoyment is low.

Keywords: Video games, entertainment, enjoyment, performance, flow, attribution theory.

1 Introduction

Many forms of contemporary entertainment computing, most importantly, video games, apply their interactivity to present tasks and challenges to users. A great diversity of task types and challenge levels can be found in modern video games, for instance, tasks that require dexterity and precise timing of control inputs (such as in first person shooters), tasks that demand logical thinking and problem solving (e.g., adventure games, brain trainers), or tasks related to strategic planning and management of complexity (such as strategy games and business simulations). While social research on the motivational appeal of video games is still sparse [1], there is some indication that resolving game tasks and mastering game challenges is closely connected to game enjoyment. Ryan, Rigby and Pryzybylski [2] demonstrated experimentally that feelings of competence are

* This research was funded by the European Commission, project “FUGA: The fun of gaming” (NEST-PATH-IMP 28765). We thankfully acknowledge the Commission‟s support.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 1–12, 2009.

an important dimension of the motivational appeal of digital games. Two surveys identified competition as driver of playing motivation [3] [4]. Both competence and competition are inherently connected to tasks and challenges presented by games – mastering challenges thus is probably linked to game enjoyment [5].

The present study addresses the issue of task resolution, mastery of game challenges, and game enjoyment in more detail. Specifically, the paper attempts to shed more light on the complexity that is involved in the connection between player performance and game enjoyment. This complexity stems primarily from the common understanding that good performance can only occur when the task resolved is not too easy. Accomplishing an easy task does not hold much merit, and performance is only valued if it is perceived as mastery of a significant challenge. Building on concepts from motivational psychology, especially attribution theory [6], this paper takes an explanatory stance on player performance and video game enjoyment by examining player responses to and enjoyment of systematically varied challenge levels.

2 Performance, Attribution, Satisfaction, and Game Enjoyment

Research in the psychology of motivation has found consistently that human individuals respond to good own performance (success) with positive emotions such as pride and joy. Weiner [6] has emphasized that such positive emotions occur if the individual identifies her-/himself as origin of the event interpreted as success. Only if the positive event can be attributed to oneself (e.g., to one's talent or one's hard effort), strong positive emotions will arise; if the individual perceives external factors (e.g., somebody else's effort or simple luck) responsible for the event, the resulting emotions may still be positive, but will not reach the same level of intensity. Conceptually, the emotion of "pride" will only occur in the case of self-attribution of the success event, whereas the emotion of "joy" may also occur in the case of external attribution of the (positive, appreciated) event. Similarly, self-determination theory [7] argues for the importance of feelings of one's own competence for positive emotions: It is thus the perception that oneself has done something 'good' or has achieved a great success that makes the difference in emotional experience.

An important underlying mechanism that connects satisfaction with one's performance to game enjoyment is self-esteem [8]. In general, increases in self-esteem go along with positive emotions such as pride and joy, and positive performance feedback or direct experience of competence rise the individual's self-esteem level [9]. Therefore, successful task resolution is theorized to level up self-esteem, and the increase of self-esteem is experienced as highly enjoyable.

Thus, in the context of video games, the challenges that players are confronted with represent opportunities to experience own competence by attributing the success events in the game to one's own skill and efforts [10]. Resolving the game tasks would then be a key to game enjoyment: As most games introduce new tasks and challenges at high frequencies, players receive ample opportunities to feel competent and successful, lift their state self-esteem, and thus generate positive self-emotions continuously throughout game play. The permanent pride of mastering ever-new game challenges would then constitute an important part of game enjoyment.

However, the link between one's own performance, self-esteem and positive emotions is more complicated, for two important moderators affect the performance-enjoyment process. One is the difficulty of the task(s) mastered, and the other is the performance expectation that the individual holds before and during task resolution. Concerning task difficulty, attribution theory [6] suggests that players cannot derive pride from the mastery of (very) easy tasks, because in this case, there is no chance to demonstrate skill or invest serious effort. Without skill demonstration and/or efforts invested, there is no reason to be proud of: The challenge is simply no challenge. In the context of video gaming, easy tasks (e.g., enemies that are easy to kill and do not cause serious damage to the player character) would rather evoke boredom than enjoyment. In contrast, (very) difficult tasks do not facilitate positive feelings either [6]. One reason is that difficult tasks are not resolved with high probability, so experiences of failure and insufficient performance arise more frequently under high difficulty conditions. Such experiences undermine self-esteem and lead to frustration and sadness – the opposite of pride, and, when applied to game situations, also the opposite of game enjoyment. A second important reason for hard tasks interrupting the effect of success on enjoyment is that if players manage to resolve very heavy game tasks, there is a often a reasonable chance that their skill and effort alone did not cause the success, but that additional external factors (e.g., luck) co-occurred, which would question the self-attribution of the success. The enjoyment value of mastering very difficult tasks is thus not as 'secure' as the fun that players can generate from mastering moderately difficult tasks. For the solution of such tasks, they can claim full responsibility, with a secure impact on positive emotions.

This consideration converges with flow theory that postulates most positive experiences resulting from mastering tasks that are not too easy and not too difficult [11]. Flow theory has also been applied to video game enjoyment [12]. Consequently, the contribution of success experiences to video game enjoyment is argued to depend on the difficulty of the game played. Neither very easy nor very hard games should elicit success-based game enjoyment to a substantial extent; rather, moderately difficult games should facilitate the highest level of enjoyment.

The link between player performance and game enjoyment is likely to be affected also by the performance expectations players hold [13]. Expert players will be convinced of their capacity to master highly difficult challenges and will thus expect themselves to be quite successful with any given new game. Novice players, in contrast, will accept the possibility of failure and underperformance in many new games. Because they are 'prepared' for failures, they should not suffer from severe reductions in game enjoyment when they face difficult tasks, as they can still meet their low performance expectations and need not be disappointed about their achievements in the game. Expert players, however, may feel frustration more frequently, as failure to accomplish (seemingly) easy or moderate tasks would imply a violation of their own high performance expectation and thus reduce their (high) self-esteem level significantly [13]. On the other hand, expert players hold an objectively better chance to master any given game task, while novice players will fail with much greater probability at any given game task. So expert players will succeed frequently but be disappointed about their few failures, whereas novice players will fail frequently but will not feel much frustration about it. It is theoretically plausible, then, to argue that good performance and achievement is virtually irrelevant for the game enjoyment of novice players, whereas for expert players, performance would be extremely important, because their self-expectations are high, and they seem to perceive video games in general as a domain to demonstrate skill and superior performance [4]. For the present study, we thus focused on expert players to examine the interplay of performance, satisfaction and game enjoyment.

From these considerations, the actual enjoyment experience during game play can be modelled. Players begin a gaming session with a self-expectation concerning their skills and performance capabilities (i.e., they define themselves as rather novice or rather expert players). The tasks that the game offers – enemies, puzzles, etc. – will lead to performance-based fun to the extent that (A) players find the tasks challenging (not too easy, not too difficult) and (B) players find their accomplishments matching their self-expectations. Because task difficulty and self-expectations are interrelated – heavy gamers find other game tasks 'hard' than novice players –, it is difficult to predict the specific level of optimal performance satisfaction that leads to increase in self-esteem, pride emotions, and thus achievement-based game enjoyment.

Moreover, recent contributions to entertainment theory [5] suggest that media users actively 'work' on their enjoyment experience, for instance, by suspending disbelief in an implausible drama plot for the purpose of maintaining a suspenseful movie experience. Because the link between player performance and game enjoyment is established through players' own evaluation of their performance (which uses perceived task difficulty and self-expectations as indicators), it is possible that players use the inherent complexity of the task-performance-satisfaction-fun connex instrumentally to preserve a maximum enjoyment even if they do not perform optimally in the game. For instance, a player who fails to kill a monster in a first-person-shooter may attribute his failure to an 'unfair' game setting that rendered the task extremely difficult or 'impossible to do'. By justifying one's own failure through external conditions (i.e., the game was unfair), the negative impact of the failure on self-esteem and the accompanying frustration and loss of fun can be buffered. So blaming the game for being unfairly difficult may be a strategy to maintain game enjoyment in spite of underperformance. Vice versa, a player who surprisingly manages to accomplish a really difficult game task may not acknowledge that luck was responsible for this success, but rather assign the great victory to his own skill, thus 'creating' a reason to be proud, with accompanying positive emotions and performance-based game enjoyment. In other words, players' evaluation of their own performance, may be biased instrumentally by players in order to maximize fun given successful game events or to preserve as much enjoyment as possible in the case of failures in gameplay. More failure may thus not necessarily lead to less enjoyment, and more success will not automatically facilitate more enjoyment.

The theoretical elaboration so far suggests that there is considerable variability in player responses to a specified game difficulty level. In order to explore the complex relationship between game difficulty, player performance and satisfaction as well as game enjoyment in more detail and to gain empirical evidence for a more accurate model of performance-based game enjoyment, the following research questions were derived.

RQ1. How does video game difficulty affect satisfaction with one's own performance in expert players?

RQ2. How does video game difficulty affect game enjoyment in expert players?

RQ3. Is the effect of game difficulty on satisfaction stronger than its effect on game enjoyment?

RQ3 explicitly addresses players' (possible) instrumental interpretation of their own performance: If players actively shape their entertainment experience, they will protect it against threats from underperformance and according frustration, and will also derive more enjoyment from mastering easy tasks than it would be appropriate from an 'objective' viewpoint (i.e., as a fair estimate of task difficulty within the attribution process would suggest). Thus, game difficulty may affect game enjoyment to a smaller degree than it affects player satisfaction: Players may not want satisfaction to dominate their fun and thus actively work against such an influence.

3 Method

To answer the research questions, an experiment with the first person shooter (FPS) "Unreal Tournament 2" © was conducted. Overall, 74 voluntary male university students aged between 18 and 32 years (M = 21.84, SD = 2.73) participated in the study. All participants said that they played at least "sometimes" computer games, and they all had at least "some" experience with FPS. Before the students were invited to the laboratory, they rated their FPS expertise on a 10-point-scale (with "1" meaning being a novice with almost no experience, "10" meaning being an absolute expert). Only individuals who rated themselves at "5" or higher were asked to participate in the study. The reason for this limitation of access was to focus on game experts (see previous section). Individuals with sufficient FPS experience were then randomly assigned to play a "duel mode" map of "Unreal Tournament 2" with either "easy", "medium" or "very hard" difficulty settings. In the "easy" condition, it was almost impossible that the player character would get hurt or died, and enemies were very easy to kill. The version with medium difficulty was supposed to provide the players with some success and the feeling of competence while a significant level of challenge was present. At last, the very difficult level was virtually impossible to win. Players necessarily got killed several times in this condition – independent from their skills. Everything aside of the difficulty level – appearance of the enemies to be dueled, map and geographical structure of the game environment etc. – was held constant across difficulty conditions. Consequently, experimental groups were confronted with systematically varying levels of game difficulty.

The participants were individually invited to a quiet room with controlled lighting conditions and were asked to sign a letter of consent to participate in the subsequent procedure. Before playing, some reaction time data were collected that are not relevant to the present analysis. Consequently, participants played their FPS level for 10 minutes. They were then requested to complete another reaction-time task and were handed a questionnaire afterwards. Players' objective performance was recorded from game statistics; for this purpose, the number of enemies killed within the 10 minute play time ("kills") and the number of times the player character was killed ("deaths") were noted by the experimenter.

Analysis of these statistics revealed that the manipulation of difficulty was highly effective (see table 1). With increasing difficulty, the average number of enemies that players managed to kill went down sharply, whereas the number of the player character's "deaths" increased substantially. These group differences were highly significant both for "kills" and "deaths".

Table 1. Average number of "kills" and own "deaths" across experimental groups of different game difficulty (n=71)

	Enemies killed		Deaths of player character	
Game difficulty	Mean	Standard Deviation	Mean	Standard Deviation
Easy (n = 25)	24	7.58	1.72	4.52
Moderate (n = 23)	8.96	5.77	15.09	5.59
very difficult (n=23)	2.65	2.81	25.74	4.85

Main effect of game difficulty on enemies killed: $F(2,68) = 86.63$, $p < .0001$; $\eta^2 = .72$.
Main effect of game difficulty on own deaths: $F(2,68) = 139.52$, $p < .0001$; $\eta^2 = .80$.

Table 2. Average ratings for game difficulty across experimental groups (n=71)

	Perceived difficulty of game	
Game difficulty (experimental factor)	Mean	Standard Deviation
Easy (n = 25)	1.58	0.61
Moderate (n = 23)	2.98	0.71
very difficult (n = 23)	3.85	0.75

Main effect of experimental variation in difficulty: $F(2,68) = 66.45$, $p < .0001$; $\eta^2 = .66$

The post-play questionnaire assessed game enjoyment (with 4 items like "the game was entertaining", scaled from "1" meaning "I do not agree at all" to "5" meaning "I fully agree", Cronbach's $\alpha = .93$), satisfaction with one's own performance (4 items like "I am proud of my performance in the game", scaled again from "1" to "5", Cronbach's $\alpha = .80$), and perceived difficulty of the game (two items on a 5 point semantic differential such as "the game was… not manageable vs. no challenge", Cronbach's $\alpha = .91$). Finally, some additional information (including demographics) was requested from participants. After responding to the questionnaire, participants were debriefed and dismissed. Each person received 5 EUR as compensation.

The postplay questions on perceived game difficulty again demonstrated the effectiveness of the experimental variation in challenge level (table 2). Players rated the difficulty level of the game level in the way the experimental manipulation had been designed; this finding also indicates that players were aware of specific objective difficulty of their game task when evaluating their performance.

4 Results

Results indicate that in general, players evaluate their performance worse if they have been confronted with harder game difficulty. Obviously, self-assessment of playing performance was made under the impression of the number of kills (and deaths) without taking the objective difficulty of the game into account. The many kills that virtually all players achieved in the easy game version seem to have caused players to rate their performance as very good in spite of the low challenge that this game version imposed. Vice versa, the many own deaths in the hardest condition obviously caused players to evaluate their performance negatively in spite of the objective difficulty level. However, overall performance ratings did not differ as strongly between players of the moderate and the hard game version, which suggests that players begin to 'defend' or 'justify' their performance when confronted with extremely difficult tasks and are not ready to accept any negative gaming outcome as consequence of their insufficient performance (table 3).

A similar analysis was conducted to examine the impact of game difficulty on game enjoyment (see table 4). Overall, the easy game version that facilitated the highest number of success experiences (enemies killed) together with the lowest number of failures (deaths of the player character) generated the most intensive game enjoyment. Enjoyment of the harder game versions was lower, with the mean difference between the moderate and the highly difficult version being smaller than the difference between the easy and the moderate condition. The greater satisfaction with one's own performance in the easier difficulty conditions thus comes along with greater enjoyment, which contradicts the assumption elaborated earlier that the moderate difficulty condition would generate more fun than the easy and the hard condition. The difference in effect sizes should be noted, as they are important for the interpretation of findings. While the effect of the manipulated game difficulty was very strong on actual performance (kills/deaths), it was a little lower but still strong ($\eta^2 = .66$) for the game difficulty rating, again substantially lower for players satisfaction ($\eta^2 = .46$), and went down to a moderate effect size for game enjoyment ($\eta^2 = .17$)

Table 3. Satisfaction with playing performance across experimental groups of different game difficulty (n=71)

Game difficulty	Mean	Standard Deviation
Easy (n = 25)	4.06	0.63
Moderate (n = 23)	2.89	0.69
very difficult (n = 23)	2.51	0.81

Main effect of game difficulty level: $F(5,65) = 28.17$, $p < .0001$; $\eta^2 = .46$

Table 4. Game enjoyment across experimental groups of different game difficulty (n=71)

Game difficulty	Mean	Standard Deviation
Easy (n = 25)	4.29	0.59
Moderate (n = 23)	3.84	0.83
very difficult (n = 23)	3.53	0.86

Main effect of game difficulty level: $F(5,65) = 6.49$, $p < .01$; $\eta^2 = .17$

5 Discussion

The experimental variation of game difficulty produced patterns of game enjoyment (RQ1 and RQ2) that are not fully in line with flow theory [12] and attribution theory of motivation [6]. From these frameworks, maximum enjoyment would be predicted for moderate game difficulty, because under such conditions, players can perceive their own skills and efforts, and attribute occurring success events to themselves. Too easy and too hard difficulty levels would either lead to boredom (or no reason to be proud on one's performance) or frustration (or the suspicion that luck helped to overcome the extreme challenges). Interestingly, our experiment found that players enjoyed the FPS the most when they were given a very easy condition with many success events (enemies killed) and very few (if any) failures (own deaths). With increasing difficulty, the number of success events went down and the number of failures events went up, and both satisfaction with the own performance and overall game enjoyment were lower than in the easy condition.

Some methodological issues need to be addressed before a conceptual discussion. First, the results may be caused by a misinterpretation of objective difficulty levels. Maybe players found the condition that the investigators labeled "easy" actually challenging, which would suggest to reconsider if the findings do in fact match with attribution and flow theory. However, players' own ratings of game difficulty were in line with the experimental manipulation: So players enjoyed the game condition the most for which they admitted a low difficulty level, which suggests that the according results are no artifact of unrealistic experimental manipulation. Second, players used the game in a laboratory setting, which may have reframed participants' situation perception in a way that made performance requirements more salient than in conventional home use situations. Reports of satisfaction and enjoyment may thus be biased due to players' intentions to impress the experimenter; however, there is no evidence for the type of bias that may have occurred. Field replications of the study may illuminate this possible problem in the future.

From a theoretical perspective, there is a need to reconcile the present results with previous studies who found evidence for flow experiences to be connected to video game enjoyment [14] and for expert players' preference for extremely difficult over simple tasks [15] – findings that seem to be just the opposite of what the present study revealed.

Our attempt to integrate these findings with previous research is focused on the issue of playing time. We suggest that the relationship between game difficulty, success rate, internal attribution of success, satisfaction with own performance, and overall game enjoyment changes over the time of using a given game. Our findings picture the situation of players beginning an unfamiliar game. In this early stage, enjoyment seems to be driven by 'quick success', that is, a high number of explicit positive feedbacks fuels game enjoyment. The fact that this fast stream of success experiences is caused by low difficulty seems to be 'ignored actively' by players: They know that the game 'makes it easy' for them, but still they have fun with being successful, although they 'objectively' did nothing to be really proud of [6]. In turn, high numbers of failure events reduce game enjoyment in the beginning stage of game use, although players know that it was an 'objectively' hard task. So our interpretation of the found link between game difficulty and game enjoyment is that during the early stage of getting acquainted to a new game, players heavily depend on visible success and positive feedback provided by the game. Internal performance evaluations such as "I was good, but the game was really hard" seem to be less important at this stage. Without such cognitive rationalization of failure, game enjoyment cannot be preserved at harder difficulty levels and goes down. Overall, players starting an unfamiliar game depend on the game's feedback of good performance; only if the game delivers such success feedback (regardless of objective difficulty), the beginning stage of game use is notably enjoyable.

These patterns of game difficulty and game enjoyment may change with increasing playing time. After eight more hours of experience with the same game, for instance, the experimental levels with easy, moderate, and difficult challenge would have different meanings for players. More importantly, players would hold more knowledge to judge their own performance independently of the game's direct success feedback. That is, with more knowledge about the game and one's own skill level, players can evaluate their performance without relying on the game's performance feedback alone. Such an experience-based interpretation of performance may then lead to a shifting pattern of performance and game enjoyment. At a later playing stage, the fact that there was no challenge involved in achieving many kills may hold more weight in the generation of game enjoyment. In turn, with more game experience, players learn to value really good performance in objectively difficult circumstances. Their experience provides the arguments to defend enjoyment even when confronted with many failures. Vice versa, their experience also enables the state of boredom when confronted with too easy tasks – the many wins that are demonstrably fun at the early playing stage become boring with increased experience.

In sum, we argue that players who are more familiar with a given game would display the pattern of game difficulty, satisfaction with own performance, and game enjoyment that is predictable from attribution theory and flow theory (see above): Game events and knowledge-based interpretation of performance jointly render moderately difficult tasks more enjoyable and create circumstances of less enjoyment for too easy and too difficult game tasks. For players who are at the beginning to using a new game, however, the pattern of game difficulty and enjoyment seems to be mostly driven by the explicit feedback given by the game, regardless of players' internal evaluations of difficulty levels. For players starting a new game, the fun of gaming thus seems to arise from what the game offers in terms of positive and negative feedback, whereas for players who are very familiar with a game, their own interpretation of the game's feedback (in terms of "how difficult was that situation?") plays a greater role in the generation of game enjoyment. This difference of patterns between early and later stages of game familiarity is thus the conceptual resolution that comes out of the present findings that seem to contradict patters reported in earlier work [6] [12].

In addition to the experimental group *mean differences* in average satisfaction and enjoyment, the *effect sizes* observed in the current experiment deserve conceptual attention. In the easy game condition, players did not 'die' very often; in the 'very hard' condition, players got killed very frequently. Effect sizes were substantially lower for players' satisfaction with their performance, and still lower for enjoyment rating, however. This means that strong group differences in objective performance led to smaller (yet massive) group differences in satisfaction with performance and to rather small (actually moderate) group differences in game enjoyment.

Concerning RQ3, this pattern of effect sizes indeed suggests that players actively manage and protect their enjoyment experience. Following arguments from general entertainment theory [5], we argue that players' strategy to maximize game enjoyment is to switch between different origins of fun instrumentally. Research on video game enjoyment has identified various sources of fun beyond 'performance – self-esteem – enjoyment', for instance, identification with the game character, or curiosity and surprise (see, for instance, [16] [17]). This means that the game experience does not fully depend on performance issues but that other factors can also affect enjoyment.

Our conclusion is thus that players do not allow performance to dominate their experience, but that they attend also to other 'fun factors' in order to maintain a positive play experience. Instrumental use of a game's fun factors would then imply to focus in performance-based enjoyment when performance is good and satisfaction is high (i.e., pride as dominating mode of game enjoyment) and to focus on other fun factors when performance is bad and satisfaction is low (i.e., curiosity or suspense as dominating mode of game enjoyment). Players seem to 'intelligently handle' the various types of fun that video games have to offer, and they seem to be able to take the most fun out of the game even if one important factor (performance) does not reach optimal values. Thus, the effect size of difficulty on enjoyment was much lower than the effect size of difficulty on actual performance. However, game difficulty and performance still have an impact on game enjoyment in spite of players' active management and protection of their fun experience (i.e., players in the high-failure condition did not report exactly the same enjoyment as players in the medium-failure condition). We thus argue that players cannot fully override game-based determinants of fun, but that they can only partly buffer the impact of fun factors on their experience.

From the perspective of game development, finally, our results provide support to the common techniques of adapting the way a game operates difficulty management to the player's current stage. For players who have just begun a new game, it is important to offer a high number of positive events (success experiences) in order to facilitate enjoyment right from the start. This can be achieved by reducing difficulty levels to the absolute minimum. Examples from successful games and their 'tutorials' for beginning players illustrate the viability of this design strategy. With ongoing game use, the difficulty level should rise stepwise in order to offer more and more information that players can use for their own individual assessment of their performance, which will contribute to sustained game enjoyment in later stages of game use. Another related strategy of game developers is to adapt the difficulty automatically and dynamically in order to adjust the game to individual performance and provide 'credible' success experiences in any stage of game exposure. The interpretation of the present results implies that such automatic difficulty adaptation should not attempt to maintain a similar level of success experiences for beginners and advanced players alike, but rather increase difficulty disproportionately fast with increasing player performance. This way, players reaching a higher game-related expertise level would be confronted with a much greater frequency of failure than early-stage players, and this change of the game's requirement profile would better fit to advanced players' expectations (because advanced players expect to run into extremely heavy game challenges and are resilient against much of the frustration that comes along with them). The present study thus suggests to compare different logics of dynamic difficulty adaptation systematically to determine their impact on game enjoyment over playing time. In general, however, the present results are nicely in line with what is widely practiced in game design concerning difficulty management and adaptation.

Finally, the findings on different effect sizes of game difficulty on player performance versus game enjoyment open interesting theoretical and applied perspectives on the video game experience. We have suggested an interpretation of the multi-causality of game enjoyment [13], which has implications for future research on user experiences in entertainment computing (i.e., to observe multiple dimensions of enjoyment simultaneously in one study and to compare expert and novice players) and for game design (i.e., orchestrate different fun factors to support players' instrumental switching between modes of fun). The overall conclusion is thus that games do facilitate fun because they are task-based environments and allow self-experiences of competence and pride, but that players are also 'smart' in construing their entertainment experience and can handle the different fun factors of video games instrumentally to maximize their emotional benefits.

References

1. Vorderer, P., Bryant, J. (eds.): Playing video games: Motives, responses, consequences. Lawrence Erlbaum Associates, Mahwah (2006)
2. Ryan, R., Rigby, C.S., Przybylski, A.: The motivational pull of video games: A self-determination theory approach. Motivation and Emotion 30, 347–363 (2006)
3. Vorderer, P., Hartmann, T., Klimmt, C.: Explaining the enjoyment of playing video games: The role of competition. In: Marinelli, D. (ed.) ICEC conference proceedings 2003: Essays on the future of interactive entertainment, pp. 107–120. Carnegie Mellon Press, Pittsburgh (2006)
4. Jansz, J., Tanis, M.: Appeal of playing online first person shooter games. Cyber Psychology and Behavior 10, 133–136 (2007)
5. Vorderer, P., Klimmt, C., Ritterfeld, U.: Enjoyment: At the heart of media entertainment. Communication Theory 14, 388–408 (2004)
6. Weiner, B.: An attribution theory of achievement motivation and emotion. Psychological Review 92, 548–573 (1985)

7. Deci, R.M., Ryan, R.L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 55, 68–78 (2000)
8. Seery, M.D., Blascovich, J., Weisbuch, M., Vick, S.B.: The relationship between self-esteem level, self-esteem stability, and cardiovascular reactions to performance feedback. Journal of Personality and Social Psychology 87, 133–145 (2004)
9. Higgins, E.T.: Self-discrepancy: A theory relating self and affect. Psychological Review 94, 319–340 (1987)
10. Klimmt, C.: Dimensions and determinants of the enjoyment of playing digital games: A three-level model. In: Copier, M., Raessens, J. (eds.) Level Up: Digital Games Research Conference, pp. 246–257. Faculty of Arts, Utrecht University, Utrecht (2003)
11. Csikszentmihalyi, M.: Flow: The psychology of optimal experience. Harper Row, New York (1990)
12. Sherry, J.L.: Flow and media enjoyment. Communication Theory 14, 328–347 (2004)
13. 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, pp. 132–145. Lawrence Erlbaum Associates, Mahwah (2006)
14. Rheinberg, F., Vollmeyer, R.: Flow-Erleben in einem Computerspiel unter experimentell variierten Bedingungen [Flow experience in a video game under experimentally varied conditions]. Zeitschrift für Psychologie 201, 161–170 (2003)
15. Behr, K.-M., Klimmt, C., Vorderer, P.: Der Zusammenhang zwischen Leistungshandeln und Unterhaltungserleben im Computerspiel [The relationship between performance and enjoyment in video games]. In: Quandt, T., Wimmer, J., Wolling, J. (eds.) Die Computerspieler – Studien zur Nutzung von Computergames, pp. 225–240. VS Verlag, Wiesbaden (2008)
16. Hefner, D., Klimmt, C., Vorderer, P.: Identification with the player character as determinant of video game enjoyment. In: Ma, L., Nakatsu, R., Rauterberg, M. (eds.) ICEC 2007. LNCS, vol. 4740, pp. 39–48. Springer, Heidelberg (2007)
17. Ravaja, N., Turpeinen, M., Saari, T., Puttonen, S., Keltikangas-Järvinen, L.: The psychophysiology of James Bond: Phasic emotional responses to violent video game events. Emotion 8, 114–120 (2008)

Analysis of Area Revisitation Patterns in World of Warcarft

Ruck Thawonmas[1], Keisuke Yoshida[1], Jing-Kai Lou[2], and Kuan-Ta Chen[2]

[1] Intelligent Computer Entertainment Lab
Graduate School of Science and Engineering, Ritsumeikan University
Kusatsu, Shiga, 525-8577, Japan
ruck@ci.ritsumei.ac.jp

[2] Multimedia Networking and Systems Lab
Institute of Information Science, Academia Sinica
Taipei 115, Taiwan
swc@iis.sinica.edu.tw

Abstract. This paper analyzes area revisitation patterns in World of Warcraft (WoW). Online-game players roam a number of in-game areas while playing the game and revisit some of them with different personal reasons. To clarify this issue, we conduct a large-scale analysis using WoW access log collected for two years consisting of more than sixty thousand characters and have discovered four main groups of area revisitation patterns. We describe also in the paper how our findings can be utilized to support both game developers and players.

1 Introduction

An online game player regularly revisits some particular areas. Their reasons differ and might be due to several factors, such as player intention, area composition and content. For example, to increase their levels, a group of players might repeatedly visit an area where monsters reside. Another area might be visited routinely because of its periodical game events. Consequently, knowledge on area groups, each with similar revisitation characteristic, has high potential in supporting of not only game developers but also players.

In this paper, inspired by a recent work on analysis of web revisitation patterns [1], we hypothesize that, similar to web pages, typical patterns exist in revisitation of online-game areas. To validate our hypothesis, we conduct a large-scale analysis using access log from World of Warcraft (WoW), a Massively Multiplayer Online Role-Playing Game (MMORPG). The contributions of this work are (a) the first analysis of this kind on an online game, (b) four area-revisitation patterns (or groups) found from access log of the world most popular online game WoW, (c) thorough discussions on each pattern, and (d) our implications to applications of the findings.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 13–23, 2009.

2 Data Acquisition and Filtering

2.1 Data Acquisition

In this work, we use access log from a Taiwanese server of WoW. At present, WoW has the highest number of subscribers [2]. WoW players must choose to belong to either Alliance or Horde faction from the beginning. Access log of Horde characters was acquired by executing /who, at our client every ten minutes during two years from January 2006 to December 2007. Before filtering, our log contains more than sixty thousand characters of all Horde races: Blood Elf, Orc, Tauren, Troll, and Undead. An execution of /who gives us information on all currently-logging-in characters and their playing areas at the execution time; other information includes the race, job, and level of the corresponding character.

2.2 Data Filtering

To increase the reliability in the data, we removed characters with total play time below 40 hours, an arguable standard RPG length. We further removed characters with at least one of the four attributes falling below 10th percentile against other characters. These four attributes are the total numbers of (a) area transitions, (b) unique areas visited, (c) and areas revisited, as well as (d) the area-transition ratio. The last one is the ratio between (a) and the log length of the character. Table 1 shows an example access log, a simplified one though, of a character. For this character, the aforementioned attributes are 6, 4, 3, and 0.6, respectively. Note that because this character consecutively visited area D, its total number of areas revisited is 3, not 4, i.e., revisiting Area A in 22 hours and 40 minutes, Area B in 50 minutes, and Area C in 22 hours and 10 minutes.

To exclude outliers, such as bots [3] and extremely hardcore players who spend an extraordinary amount of time in the game, we further filtered out characters

Table 1. Example of simplified access log for explanation of four character attrbutes

Time	Area
10/08/06 02:12:37	A
10/08/06 02:22:28	A
10/08/06 02:32:37	B
10/08/06 02:42:38	C
10/08/06 03:02:30	D
10/08/06 03:12:37	D
10/08/06 03:22:35	B
10/09/06 00:52:37	C
10/09/06 01:02:37	A
10/12/06 23:36:37	A

with log length of 99th percentile against other characters. As a result, the number of remaining characters becomes 6491. For area filtering, we removed areas with the integral of the revisitation curve, defined in the next section, of 1st percentile against other areas. This results in 107 areas remaining after filtering.

3 Analysis

Our analysis approach follows the recipe in [1]. First, we use the revisitation curve to characterize an area of interest. Then we cluster these areas into related groups using a hierarchical clustering algorithm.

The revisitation curve of an area represents the total number of times the area is revisited by characters within in a given time interval. Thirteen exponential time bins are employed, i.e., within 32, 64, 98, 136, 212, 424, 848, 1696, 3392, 6784, 13568, 27136 mins, and above; representing approximately, 30 min, 1 hour, 1.5 hour, 2 hours, 3 hours, 6 hours, 12 hours, 1 day, 2 days, 4 days, 8 days, 16 days, and above, respectively. For the example log in Table 1, the data of this character will add 1 to the 2nd bin of Area B and to the 8th bin of Areas A as well as C. To facilitate comparisons between areas, normalization is performed for the revisitation curve of area$[i]$ as follows:

$$\frac{\text{bin}[j] \text{ of area}[i]}{\text{average of bin}[j] \text{ among all areas}}$$

For area$[i]$, each of the thirteen normalized bin values represents an element of 13-dimensional input vector i used in cluster analysis described below.

To cluster areas into multiple groups, each having similar revisitation characteristic, we adopt commonly used Ward hierarchical clustering [4], available in R. Because shape information is important, the cosine distance is selected in Ward clustering. The index in use for automatically deciding the number of clusters is maximization of

$$\frac{\text{interclass variance}/(\text{number of clusters} - 1)}{\text{intraclass variance}/(\text{number of areas} - \text{number of clusters})}$$

[5]. Because we anticipate that there should also be four main groups as in [1], the minimum number of clusters is set to four.

4 Results and Discussions

Figure 1 shows the clustering results where all areas are divided into 15 clusters. Each cluster is labeled to one of the four groups, i.e., Fast, Medium, Slow, and Hybrid, based on the revisitation curve's shape of the cluster centroid. Table 2 summarizes the cluster information for Fast, Medium, and Slow; and Table 3

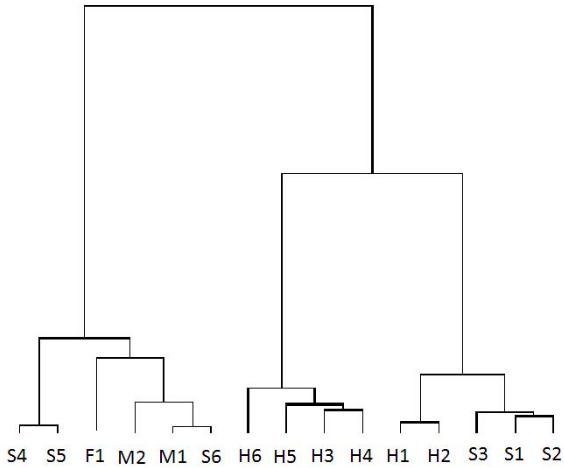


Fig. 1. Hierarchical clustering results

summarizes the cluster information for Hybrid. Each group is discussed in detail in the following.

4.1 Fast

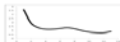
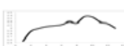
Areas in this group are often revisited within a short interval, say, less than hour. The areas in this group can be classified into two categories: battlegrounds and starting areas.

4.1.1 Battleground

One of the main features of the Warcraft universe is the raging conflict between Horde and the Alliance. Battlegrounds are instanced areas solely created for PvP (player-versus-player) combats. The two initial Battleground areas, Alterac Valley and Warsong Gulch, went online on June 7th, 2005. Arathi Basin was introduced in Patch 1.7 on September 13th, 2005. The fourth battleground, named Eye of the Storm, was released with the Burning Crusade extension pack.

Battlegrounds are instances and have a limit on the number of players who can enter at the same time, where the limit varies in different battlegrounds. The game provides Battleground Queuing System to assist players register for the join battlegrounds given the player number limits. Players can wait for entering a battleground anywhere in the world via the game's user interface. Once it is the turn for a player to enter the battleground, he will be teleported into the battleground, and teleported back after the battle finishes. All the battlegrounds are classified as fast-revisiting areas, which indicates that a battle is usually finished in 30 minutes and PvP players often reenter a battleground immediately for the next match.

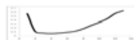
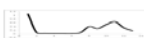
Table 2. Summary of cluster information for Fast, Medium, and Slow

Group	Cluster	Centroid Curve	Number of Areas	Area Names
Fast	F1		7	Alterac Valley, Arathi Basin, Eversong Woods, Eye of the Storm, Ghostlands, Maraudon, Warsong Gulch
Medium	M1		8	Blade's Edge Mountains, Hellfire Peninsula, Nagrand, Netherstorm, Shadowmoon Valley, Terokkar Forest, Undercity, Zangarmarsh
	M2		2	Orgrimmar, Shattrath City
Slow	S1		6	Duskwood, Elwynn Forest, Loch Modan, Redridge Mountains, Twisting Nether, Wetlands
	S2		5	Blackfathom Deeps, Ragefire Chasm, Razorfen Downs, Razorfen Kraul, Shadowfang Keep
	S3		11	Blasted Lands, Burning Steppes, Darkshore, Dustwallow Marsh, Hall of Legends, Moonglade, Mulgore, Scarlet Monastery, Swamp of Sorrows, The Temple of Atal'Hakkar, Wailing Caverns
	S4		14	Alterac Mountains, Arathi Highlands, Ashenvale, Azshara, Badlands, Desolace, Feralas, Hillsbrad Foothills, Silverpine Forest, Tanaris, The Hinterlands, Thousand Needles, Thunder Bluff, Un'Goro Crater
	S5		6	Blackrock Mountain, Deadwind Pass, Durotar, Gates of Ahn'Qiraj, Searing Gorge, Tirisfal Glades
	S6		9	Eastern Plaguelands, Felwood, Silithus, Silvermoon City, Stonetalon Mountains, Stranglethorn Vale, The Barrens, Western Plaguelands, Winterspring

4.1.2 Starting Area

During our data collection period (Jan 2006 to Dec 2007), an extension pack called "The Burning Crusade" was released in March 2007. In this extension pack, there are two new races introduced: The Draenei for the Alliance and the Blood Elves for the Horde. New Blood Elf characters start their journey in the Eversong Woods, the northern section of Quel'Thalas. Farther south is the Ghostlands, a level 10-20 zone. Low-level Blood Elves gain experience points and reward items by pursuing the quests in Eversong Woods and Ghostlands. However, most of the quest givers are in the capital city, Silvermoon (S6), which

Table 3. Summary of cluster information for Hybrid

Group	Cluster	Centroid Curve	Number of Areas	Area Names
Hybrid	H1		11	Dire Maul, Hellfire Ramparts, Scholomance, Shadow Labyrinth, Stratholme, The Arcatraz, The Black Morass, The Botanica, The Shattered Halls, The Steamvault, The Underbog
	H2		10	Auchenai Crypts, Blackrock Depths, Gnomeregan, Mana-Tombs, Old Hillsbrad Foothills, Sethekk Halls, The Blood Furnace, The Slave Pens, Uldaman, Zul'Farrak
	H3		5	Blackrock Spire, Naxxramas, Ruins of Ahn'Qiraj, The Mechanar, Zul'Gurub
	H4		3	Blade's Edge Arena, Nagrand Arena, Ruins of Lordaeron
	H5		5	Gruul's Lair, Magtheridon's Lair, Molten Core, Onyxia's Lair, Tempest Keep
	H6		5	Ahn'Qiraj, Blackwing Lair, Karazhan, Serpentshrine Cavern, Zul'Aman

is located in the northeastern part of the Eversong Woods. Therefore, low-level Blood Elves have to regularly return to Silvermoon to complete their quests or to take new quests. After that, they tend to go back to the starting areas and carry on their new quests. The fast revisitation of starting areas confirms that the quests for low-level players are relative easy and can be done within a short time.

4.2 Medium

Areas in this cluster have a peak in the left middle of the curve, i.e., between 1 hour and 1 day. We find that the areas in this cluster are due to two possible reasons as follows.

4.2.1 Daily Quest

Daily quests (often called "dailies") are repeatable quests firstly introduced in Patch 2.1. Areas for daily quests are included in M1, such as etherstorm and Shadowmoon Valley. These quests provide significantly better rewards than regular quests. A player can complete up to 25 daily quests in each day. After a player solves 25 daily requests, he must wait for the "daily quest reset" event. The exact time for such reset depends on the configuration of each realm, which is often chosen in the early morning. Thus, it is reasonable for players to revisit the areas which provide daily quests between 1 hour and 1 day.

4.2.2 Bank and Auction House

Other areas in the Medium cluster are capital cities, including Undercity (M1) for the Undead, Orgrimmar (M2) for the Orc and Troll, and Shattrath city (M2) for the united force. The cities serve as the main cultural, political, and economical center of each race. To players, one of the main functionalities of the capital cities is their economical functions, i.e., banking and auction. Banks and auction houses are usually located only in major cities or towns, especially in capital cities. A bank is a building with teller NPCs which allow a player to access his own private storage. A player can freely deposit and withdraw his items and currency in the storage via a bank or its branches. An auction house is a place for players to trade goods, weapons, and equipments. As of the release time of Patch 1.9, each capital city has its own auction house. The medium revisitation patterns of capital cities implies that players tend to regularly come back to their respective capital cities to access their private storage or trade goods with others.

4.3 Slow

The revisitation curves in this cluster have a slow peak, i.e., revisitation intervals are longer than 1 day. The level 30-70 areas are included. The revisitation curves of growth areas for moderate and upper level (30-70) characters, included in S4, S5, S6, have higher amplitudes than those of lower-level growth areas.

4.3.1 Low-Level Instance

Those instances are special areas where a group of players can interact privately within a dungeon. Instance dungeons tend to feature the most difficult and rewarding content. Since the majority of the players had achieved the 70th level in our traces, the low-level (20-30 level) instances, such as Blackfathom Deeps, Ragefire Chasm, Razorfen Downs, Razorfen Kraul, and Shadowfang Keep (all in S2), can no longer offer satisfying rewards. Such low-level instances thus become less popular, and therefore, players do not revisit the areas frequently.

4.3.2 Monthly Event

Another reason for slow revisitations is due to a monthly event, the Darkmoon Faire (the Faire), which is a regular event firstly introduced in Patch 1.6. The Faires always start on the first Monday of a month, and the location rotates among three places, namely, Elwynn Forest (S1), Mulgore (S3), and Terokkar Forest (M1). The event lasts for one week and appears in turn in the three places each month. Because some special goods can only be purchased in the Fairs, players may manage to attend such events, even the location is far away from their homelands. For example, when the Faire takes place in the Elwynn Forest near the Alliance land, we find that some Horde players traveled through Dust Wood, Loch Modan, Redridge Mountains, Twisting Nether, and Wetland (all in S1) in order to get to Elwynn Forest. Because these in-between areas are unpopular and may be only visited due to such rare events, those areas exhibit slow revisitation patterns and are included in the slow revisitation cluster.

4.4 Hybrid

4.4.1 High-Level Instance

High-level instances feature the most difficult and rewarding content, thus such instances are very popular. For example, Stratholme, Old Hillsbrad Foothills, Blackrock Spire, Molten Core, and Ahn'Qiraj which associate with H1, H2, H3, H5, and H6, respectively, are in this category. Due to the challenge in solving quests offered in these areas, very often the entire party of players are eliminated by the enemies and transferred to the nearest land outside the instance. This kind of situation is called the "wipeout". After a wipeout, players usually gather and enter the same instance within 10 minutes, which causes the left peak in the revisitation curve.

On the other hand, in order to control the number of valuable items from inflating, an instance may have a "lockout" period after it is conquered. During the lockout period, the players who completed the instance are not allowed to re-enter the instance. The lockout period is often set to one day or one week depending on the scarcity of the reward. This design leads to the right peak in the revisitation curves for these areas.

4.4.2 Arena

The remaining areas in the Hybrid cluster are mostly arenas, including Blade's Edege Arena, Nagrand Arena, and Ruins of Lordaeron (all in H4). An arena is a place in which teams of players compete against each other to gain the Arena Points in deathmatch-like player-vs-player combats. Since such matches often conclude in 10 minutes, players tend to reenter an arena frequently. This is the reason for the left peak.

However, the Arena Points are not awarded after each match. For example, one requirement to gain the Arena Points is to attend at least ten matches within a week, and the Points is awarded on Thursdays regularly. Therefore, players tend to reenter an arena to gain rewarding points weekly. We consider this design the cause to the right peak in the revisitation curve.

5 Design Implications

Design implications are given in this section. Although WoW is used in the analysis in this paper, we note here that our implications are applicable to other MMORPGs.

5.1 Area Design Reconsideration

Through the analysis of players' area revisitation patterns, game designers can check whether the usage of areas conforms to their expectation. Specifically, some areas may be unreasonably popular where players revisit them very frequently, while some others may be unexpectedly unpopular where players seldom revisit. The unexpectedly popular areas might be because players have found some tricks that can defeat "boss" monsters or gain a great quantity of reward, whereas the unexpectedly unpopular areas might be because the game missions provided are too difficult or the reward is not worth enough. Thus, we consider area revisitation patterns could be a useful summarization tool of game players' passion about certain areas, and, at the same time, a useful analytical tool for game designers to reconsider the design of game areas.

In addition, to maintain high user satisfaction, the contents of Fast and Medium areas should be updated with higher frequencies than those of Slow. Fast and Medium areas are more popular ones where players more frequently visit. Neglect in updating them may lead to player retirement.

5.2 Players' Social Network Solidification

For a group of players, if their characters share similar area revisitation patterns, it is likely that they share similar game play preference and strategies, and their characters have same game levels and expertise. For example, a group of players may all prefer to boost the game levels through slashing monsters in caves rather than simply chatting with other players in towns. On the other hand, some other players may prefer accumulating valuables by manufacturing goods, provide services, or trading, thus they tend to spend a lot of time in market places and rarely get into battlefield areas. Via the clustering of area revisitation patterns we are able to perform player clustering, where players in the same cluster share similar game play preference and interests.

Game operators can exploit this kind of information to solidify the social network between gamers by designing systems like "friend recommendation" or by holding social events that encourage players to know about each other. For players who share similar game play strategies, they are more likely to connect to and share with each other, which will establish social relationships between those players and further solidify the overall social network of the game's participants. Consequently, this will increase the stickiness to the game and furthermore prolong the game's lifetime.

5.3 Player Support

Individual player support can be provided based on area revisitation characteristic of a player of interest. For example, assume that the revisitation curves of a character of interest to battle areas exhibit Slow, but those to growth areas are Fast. This indicates that the player does not favor fighting against other players, but rather prefers to raise his/her character's level. The game system can thus give to this player personal supports such as a direct invitation to an event at a growth area. In addition, the game system can provide to players a kind of reminder service that reminds a player of interest to visit an area whose blank interval since the last visit has passed the peak time in the revistation curve.

If an area search function is available in the game, information on area revisitation can be used in area retrieval computation. For example, assigning more preferences to Slow areas, than to Fast or Medium areas, in the search result might give players higher satisfaction. This is because we think there is a higher possibility that players use the area search function to locate less known areas, such as those in Slow group.

6 Related Work

Since an early report [6] on a large amount of re-visit information in web navigation, many studies [7-11] have been conducted to understand webpage revisitation and browsing behavior. Compared to these studies, the work in [1] that we base on is the largest study in terms of the number of participants. Very recently, the same group of authors has furthered their work and summarized their recent findings on the association between change in content and revisitation in [12].

WoW has been used as a research platform in many studies because of its popularity and availability of WoW API[13], allowing interaction and modification of the game. In [14] and [15], automatically collected data were explored to understand gaming experiences and social dynamics, respectively. The predictability of online-game players' subscription time and the effect of changes in game features to social interaction were studied in [16] and [17], respectively. Very recently, a system for automatically generating comics from WoW gaming experiences has been developed [18].

7 Conclusions

Inspired by an existing work on analysis of web revisitation patterns, in this paper, we analyzed area revisitation patterns in WoW. As in web, areas were divided into four groups Fast, Medium, Slow and Hybrid according to their revisitation curves. For each group, we highlighted representative areas and their characteristics associated with their curves. A number of design implications were given that support not only the game-developer side but also the game-player side. Extensive tests of these implications are left as our future work.

Acknowledgments

This work was supported in part by Japan Society for Promotion of Science (JSPS) under the Grant-in-Aid for Scientific Research (C) 20500146 as well as by the National Science Council under the grant NSC97-2221-E-001-009 (Taiwan).

References

1. Adar, E., Teevan, J., Dumais, S.: Large Scale Analysis of Web Revisitation Patterns. In: Proc. of the twenty-sixth annual SIGCHI conference on Human factors in computing systems (CHI 2008), pp. 1197–1206 (2008)
2. http://www.mmogchart.com/Chart1.html
3. Thawonmas, R., Kashifuji, Y., Chen, K.T.: Detection of MMORPG Bots Based on Behavior Analysis. In: Proc. of ACM SIGCHI International Conference on Advances in Entertainment Technology (ACE 2008), pp. 91–94 (2008)
4. Ward, J.H.: Hierarchical Grouping to optimize an objective function. Journal of American Statistical Association 58(301), 236–244 (1963)

5. Calinski, T., Harabasz, J.: A dendrite method for cluster analysis. Communications in statistics 3(1), 1–27 (1974)
6. Catledge, L.D., Pitkow, J.E.: Characterizing browsing strategies in the World-Wide Web. In: Proc. of the Third International World-Wide Web conference on Technology, tools and applications (WWW 1995), pp. 1065–1073 (1995)
7. Tauscher, L., Greenberg, S.: How people revisit Web pages: Empirical findings and implications for the design of history systems. Int. J. of Human-Computer Studies 47(1), 97–137 (1997)
8. Cockburn, A., McKenzie, B.: What do Web users do? An empirical analysis of Web use. Int. J. of Human-Computer Studies 54(6), 903–922 (2001)
9. Herder, E.: Characterizations of user Web revisit behavior. In: Proc. of Workshop on Adaptivity and User Modeling in Interactive Systems (ABIS 2005), pp. 32–37 (2005)
10. Kellar, M., Watters, C., Shepherd, M.: A goal-based classification of Web information tasks. In: Proceedings of the Annual Meeting of the American Society for Information Science and Technology (ASIS&T 2006) (CD-ROM Proceedings) (2006)
11. Obendorf, H., Weinreich, H., Herder, E., Mayer, M.: Web page revisitation revisited: Implications of a long-term click-stream study of a Long-term Click-stream Study of Browser Usage. In: Proc. of the twenty-fifth annual SIGCHI conference on Human factors in computing systems (CHI 2007), pp. 597–606 (2007)
12. Adar, E., Teevan, J., Dumais, S.: Resonance on the Web: Web Dynamics and Revisitation Patterns. In: Proc. of the twenty-seventh annual SIGCHI conference on Human factors in computing systems (CHI 2009), pp. 1381–1390 (2009)
13. http://www.wowwiki.com/World_of_Warcraft_API
14. Ducheneaut, N., Yee, N., Nickell, E., Moore, R.J.: Building a MMO with mass appeal: a look at gameplay in World of Warcraft. Games and Culture 1(4), 281–317 (2006)
15. Williams, D., Ducheneaut, N., Xiong, L., Zhang, Y., Yee, N., Nickell, E.: From tree house to barracks: The social life of guilds in World of Warcraft. Games and Culture 1(4), 338–361 (2006)
16. Tarng, P.-Y., Chen, K.T., Huang, P.: An Analysis of WoW Players' Game Hours. In: Proc. of the seventh annual Workshop on Network and Systems Support for Games, NetGames 2008 (2008)
17. Chen, V., Duh, H., Hong, R.: The Changing Dynamic of Social Interaction in World of Warcraft: The Impacts of Game Feature Change. In: Proc. of ACM SIGCHI International Conference on Advances in Entertainment Technology (ACE 2008), pp. 356–359 (2008)
18. 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)

Scaling the Level of Difficulty in Single Player Video Games

Maria-Virginia Aponte, Guillaume Levieux, and Stéphane Natkin

Centre d'Etudes et de Recherches en Informatique du CNAM
Conservatoire National des Arts et Métiers, Paris, France
forename.lastname@cnam.fr

Abstract. In this this paper, we discuss the interest and the need to evaluate the difficulty of single player video games. We first show the importance of difficulty, drawing from semiotics to explain the link between tension-resolution cycles, and challenge with the player's enjoyment. Then, we report related work on automatic gameplay analysis. We show through a simple experimentation that automatic video game analysis is both practicable and can lead to interesting results. We argue that automatic analysis tools are limited if they do not consider difficulty from the player point of view. The last section provides a player and Game Design oriented definition of the challenge and difficulty notions in games. As a consequence we derive the property that must fulfill a measurable definition of difficulty.

Keywords : video games, challenge, difficulty, learning, evaluation.

1 Introduction

One of the fundamental issues to tackle in the design of video games is mostly referred as *creating a well-shaped difficulty curve*. This means that one of the core element of a good game design is to make the game just as difficult as it has to be, so that the player feels challenged enough, but not too much. However, game creators cannot rely on strong tools to help them in this task, and there is not even a clear and accepted definition of difficulty as a measurable parameter. For now, game difficulty adjustment is a subjective and iterative process. Level / game designers create a sequence of challenges and set their parameters to match their chosen difficulty curve. Finding the right sequence and tune every challenge relies on playtesting performed by the designers, and, at some important milestones, by focus test groups. Focus tests are costly and it's very hard for a designer to evaluate the difficulty of a challenge he created and played for many hours. Our goal is to provide a clear, general and measurable definition of the difficulty in games. We must rely on accepted definitions of video games and works which relates the difficulty in games to the quality of the games, as perceived by the player. We present related work on automatic gameplay analysis, and then report a first experiment with a basic synthetic player. Finally, we define difficulty, tacking into account the player experience.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 24–35, 2009.

2 Scaling the Difficulty

Difficulty scaling is a fundamental part of game design [1] [2]. However, this is not an obvious consequence of accepted definitions of video game. Jesper Juul has listed many of them and has proposed a synthesis [3]. We start from Juul's definition to explain why difficulty scaling is so important in game design :

> *'A game is a rule-based formal system with a variable and quantifiable outcome, where different outcomes are assigned different values, the player exerts effort in order to influence the outcome, the player feels attached to the outcome, and the consequences of the activity are optional and negotiable.'*

This definition gives a clear, precise idea of how a game system behaves, and manages to take into account the most interesting parts of the previous definitions. But for our purpose, we must explain more precisely why difficulty is such an important thing. The fact that the player *exerts effort in order to influence the outcome*, and *feels attached to the outcome* is the core point. To point out the important components of a gameplay, and foremost the link between caring about difficulty and making a good game, it is necessary to coin a definition that leaves aside the game's dynamics structure and focuses on video games from the player's point of view.

Robin Hunicke describes a game using a *Mechanics, Dynamics and Aesthetics (MDA)* framework [4]. Mechanics are the tools we use to build a game (e.g. physics engines, pathfinding algorithm...), Dynamics describes the way the Mechanic's components behave in response to the player, and Aesthetics is the desirable emotional responses evoked to the player. Of course, the design goals is the Aesthetics, that is to say the player's emotions. We argue that the difficulty of challenges greatly influences video game's aesthetics and thus play a central role in game design.

Umberto Eco's book *The open work* is a fundamental research about interactive art's aesthetics [5]. Umberto Eco states that when we face a piece of art, we are interpreting it, seeking patterns, looking for information. Depending on our culture and knowledge, we will find something to grab on within the stimulating field of the piece of art. But then we will go further, and find another interpretation and feel lost for short moment, while shaping our new pattern. Moreover, when a piece of art is interactive, the aesthetic value comes both from the tension resolution and from the fact that this resolution is a consequence of our choice. Assuming that a video game is an open work we can propose a similar analysis. Every time the player faces an obstacle, he gets lost for a few seconds. Then he finds and choose a pattern, press the right buttons, and takes pleasure both from resolving a tension and from making a choice. Thus, we can draw from Umberto Eco's work that in video games, challenge is fundamental because it creates tension situations that the player has to solve and the opportunity of meaningful choices.

Related work on video game player's enjoyment support our analysis and place challenge at the center of video game's aesthetics. In his book *A Theory of Fun* *for Game Design*, Ralph Koster states that we have fun playing games when we discover a new pattern, i.e. a strategy that we apply to overcome a challenge [6]. Sweetser and al see challenge as one of the most important part of their *Game Flow* framework [7]. Yannakakis et al measure player enjoyment from challenge, besides behavior and spatial diversity [8].

Mihaly Csikszentmihalyi's Theory of Flow, that researchers have applied to video game as a measure of the player's enjoyment, helps us to make a link between the difficulty of a challenge and the player's enjoyment [9] [10] [7]. A player is in a *Flow* status, and thus enjoying the game, when the task is neither too hard nor too easy. It is thus not enough to create tensions situations and to give the player choices to resolve this tension, a good game design must accurately scale the difficulty of a challenge to have a tension level that leads to the player's enjoyment. Thus, a definition of a game from the Aesthetic point of view and centered on challenges could be:

> *'Regarding challenges, the Aesthetics of a game is created by tension-resolution cycles, where the tension is kept under a certain threshold, and where the resolution of a cycle depends on the player's choices.'*

This definition doesn't take into account every aspect of game aesthetic but is focused on challenge, that most studies consider as a core component of game's aesthetics. Tension situations that the player seeks and try to solve have been created by the game designer and the amount of tension they deliver directly stems from their complexity. As a result, difficulty scaling is a central task of a good game design. Games already propose different difficulty levels [1], and sometimes even Dynamic Difficulty Adjustment [2], manipulating some specific parameters of the gameplay in real time [4], or automatically scaling the game AI capacity [11]. But whichever difficulty scaling method the game designer uses, he must still tune them properly. It is sometimes really hard to guess to which extent a change in a low level parameter will just make the game a bit harder or dramatically change the gameplay [1], and tuning is one of the most time consuming area in game AI development [12]. This is this design process that we want to shorten by providing tools that will help game designers evaluating the impact of any difficulty scaling parameter on the final difficulty curve. To create good gameplay, it's then fundamental to provide game designers with strong tools and a definition of difficulty as a measurable parameter.

3 Related Work: Testing with a Synthetic Player

Our goal is to evaluate a parameter or a set of parameters that can be considered as a measure of a game difficulty. There are two theoretical approaches to evaluate such a parameter. The first way is to find, according to the game structure, a mathematical expression of the parameter and to solve the corresponding equations. The complexity of a game and the notion of difficulty tends to show that this approach is not practicable. A second solution is to experiment the game and measure the parameter. To experiment the game we may either use a real or a synthetic player. The main advantage of a real player is that he behaves like a real player. In counterpart he plays slowly, becomes tired and his behavior is only known through the game interface. The synthetic player is tireless, plays quickly and his behavior can be fully understood. The design of the synthetic player allows to simulate some foreseen behavior of a real player (risky or careful, for example) and some simple learning techniques.

Gameplay testing has already been the subject of many interesting researches. Alasdair Macleod studied gameplay testing of Perudo, a bidding dice game, simulating plays with a multi-agent system [13]. He wanted to modify Perudo's gameplay to make it more fair, and added a rule he thought it would help loosing players to stay in the game. Running the experiment and analyzing the modified game, he obtained the counter-intuitive result that the rule was not helping loosing players at all. These results shows that self-play testing can help testing gameplay modifications.

Neil Kirby analyzed Minesweeper, replacing the player by a rule based AI [14]. Each rule was related to a different play complexity. He found out that Minesweeper was surprisingly not as hard as he supposed it to be, as the most part of the board was often solved using only the very simple rule. These results point out that automated techniques can provide interesting approaches to study video game difficulty.

Both Perudo and Minesweeper are simple games, but automated analysis can also be applied to complex off-the-shelf games. Bullen et al used Unreal Engine (Epic Games) and created a gameplay mutator providing sensors to log useful game events [15]. They tested Unreal Tournament 2004 (Epic Games) using partial and fully automated testing (i.e. both during player vs AI and only AI games). They pointed out that fully automated tests had to be done with a specific AI, because standard AI was not aggressive enough. The fact is that standard Unreal Tournament AI has been created to entertain the player, not to mimic his behavior, and thus is not able to fully explore the gameplay. Recently, Lankveld et al proposed to analyze a game difficulty using incongruity, the distance between the actual dynamics of the game and the mental model the player has built [16]. They plan to infer the complexity of the player's mental model, and thus the difficulty of the game, by monitoring his actions. These works show that, to be useful, a synthetic player must simulate in some way a real player.

Automated game analysis can be done at several levels. Nantes et al distinguish Entertainment Inspection (i.e. gameplay testing), Environment Integrity Inspection (i.e. Sounds, graphics related issues) and Software Inspection [17]. Their system targets Environment Integrity Inspection, using Computer Vision, and especially corner detection to detect aliasing issues in shadows rendering. This is a complementary approach to the one we propose, and Nantes et al acknowledge the need of analysis tools at every inspection level.

As we argued in the previous section, Machine Learning is particularly interesting in automated gameplay testing. If we want the synthetic player to test behaviors we didn't think about before, then it must explore the game state space by himself. Many researchers explore how machine learning can be helpful to video game development, and especially concerning automated testing. Chan et al used a Genetic Algorithm to create sequences of actions corresponding to unwanted behavior in FIFA-99 (EA Games) [18]. They also consider that the game dynamics is too complex too be fully formalized, because of huge branching factor, indeterminism and the fact that even designers never formally define it. There is thus a need to build an AI driven agent to explore this dynamics, here using evolution techniques. Spronck et al also took the same approach, making neural networks evolve to test a simplified version of the spaceships game PICOVERSE, providing an insightful analysis of its AI driven opponents [19].

Automated learning approaches becomes inadequate when it comes to creating characters with complex behaviors automatically from scratch, as sated John E. Laird [20]. But many researchers use games to evaluate machine learning techniques. The game is often considered as a reference problem to be solved by the AI. Pacman (Namco) [21], for example, has been the subject of many researches, applying various techniques to create synthetic players, from Neural Network Evolution [22] [23] [24], to Reinforcement Learning [25], Genetic Programming [26] and genetic evolution of a rule-based system [27]. Yannakakis et al [8] takes another point of view. They use synthetic characters to maximize player enjoyment, and validate their measure of enjoyment, based on challenge, behavior diversity and spatial diversity . These results show that machine learning techniques can be useful when analyzing a gameplay.

4 Case Study

4.1 The Experiment

These sections present an analysis of a Pacman-like predator-prey computer game. We built a simplified version of Pacman, using only one ghost chasing Pacman. Both Pacman and the ghost use A* pathfinding, the whole graph and shortest path being built at startup and stored to save calculation power. The challenge is to eat a maximal number of pellet without being killed by the ghost. The synthetic player has a partial view of the game. It knows five parameters. The first one had four values, giving the direction of the nearest pellet. The four other dimensions describe the four Pacman directions. For each direction, Pacman knows whether he is facing a wall (0), a ghost one (1) or two (2) step away from him, a free-of-ghosts path less than 18 steps long (3), or free-of-ghosts path longer than 18 steps long (4). We choose this game state abstraction because we consider that the main information that a real player uses is the path to the nearest pellet and the safety of the fourth direction he can take.

The only ghost in the maze, Blinky, has been programmed to chase the player, taking the shortest path to reach him. In Pacman original rules, ghosts periodically enter *scatter mode* and stop chasing the player to go back to their respective board corner. But as a first step, we wanted to maintain the rules at their minimum complexity, so that results could be more easily interpreted. The synthetic player AI was programmed with a Markov Decision Process, using reinforcement learning with Q-Learning algorithm with eligibility traces ($Q(\lambda)$) [28]. Parameters for $Q(\lambda)$ were $\alpha = 0.1$, $\gamma = 0.95$, $\lambda = 0.90$. We balanced exploration and exploitation using ϵ-greedy action selection algorithm, with $\epsilon = 0.05$.

We consider the analysis of Pacman difficulty according to a single gameplay parameter: the player speed. The number of pellets eaten by Pacman is our difficulty evaluation function. We choose this parameter because in Pacman original gameplay, ghosts / Pacman relative speed is already used to scale difficulty [21]. Every 14 frame, Blinky changes its position, moving one step up, down, left or right. Each step is 16 pixel long, the distance between two pellets. Besides going up, down, left or right like the ghost, the player also has the option to do nothing and stay at the same place. We tested the synthetic player's performance for different gameplay configuration.

4.2 Results

We run six experiments with speed varying from 0 (i.e. Pacman and the ghost move every 14 frame) to 7 (i.e. Blinky still moves every 14 frames but Pacman moves every 7 frame). We let the synthetic player develop his policy during 50000 games, Pacman having 3 lives per game.

The figure 1 presents a synthetic view of these results. What we can extrapolate from these is that modifying Pacman's speed, the difficulty tends not to be modified in a linear way. There is much less difference in Pacman score between speed 0 and 5 than between speeds 5 and 7. Such an analysis could be useful for a game designer when tuning the difficulty. He can understand that when Pacman speed gets closer to twice the ghost speed, then the games get really easier. Between 0 and 5, difficulty raises almost linearly.

4.3 Critical Analysis

These results show that it is possible to evaluate a difficulty curve, for a given game with a given challenge whose difficulty can be tuned according to a given parameter. However, this experiment is just an attempt to describe difficulty for a specific game. It doesn't give us a general framework we could apply to any game to measure its difficulty. The next step is thus to find a general and precise definition of the difficulty.

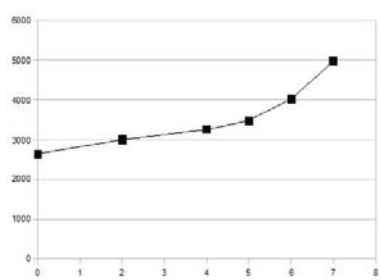


Fig. 1. Pacman score at different speeds - 5000 last games mean value

5 The Meaning of Challenge and Difficulty

To this point we have used the term of difficulty in games without providing any definition of this word. This is a difficult point. First, one cannot talk about difficulty without referring to the ability to perform an action in a given situation. We shall not try to give a general definition of difficulty covering a wide range of psychological aspects from emotional problems to intellectual and physical challenges. We consider the notion of difficulty in the sense used in Game Design.

5.1 Challenges

In all these cases, a level of difficulty is related to the player skills or abilities to overcome a given challenge. We must first define the notion of challenge in games. Starting from Juul's definition, we can consider that a video game challenge is by itself a sub-game: a rule based system with variable and quantified outcomes. According to the quantification, some of the outcomes may be considered either as a success or a failure. A general definition of the difficulty has to take into account that the quantification can be binary (WIN, LOSE) or discrete (from 0 to N points). But the definition must not depend on the units chosen, only on the relation between the outcome value and the feeling of victory or defeat. This relation is an important design choice and is explicitly implemented as a feedback in the game. Thus, we consider that in all cases the designer can define a binary function that can decide whether the player has won or lost.

The notion of difficulty is also related to time. At a given time a player may be still trying to overcome the challenge or has either won or lost. The time to overcome the challenge is also related to its difficulty: one may take hours to solve a difficult chess problem and a few minutes to solve a simple one. But the time to overcome a challenge is also relative to the conditions of the challenge itself. The player has a few seconds to choose an attack in a fighting game, a few minutes to choose a move in a RTS (Real Time Strategy) game and an unbounded delay to take a decision in a turn by turn game. In certain games the time is a criteria of evaluation: you have five minutes to find a bomb before explosion, your rank in a ski slalom game (as in WII FIT) depends on the time taken to finish the slalom. But in all cases, we can assume that the designer knows the minimal time needed to decide whether the challenge is overcome or not.

This leads to the two following definitions: *A video game challenge is a dynamic rule based system with two outcomes WIN or LOSE. At a given time t a challenge can be in one of the four possible states NOT STARTED, IN PROGESS, WIN, LOSE, according to the following automaton (Fig. 2).*

A solution of a challenge is a sequence of player's actions and the corresponding game feedback that leads the automaton from the NOT STARTED state to the WIN state.

In games solving a challenge may imply to solve a set of sub challenges. The structure of the game as a set of quests and levels is a logical, topological and, as a consequence, temporal combination of challenges (see [29] for a formal definition

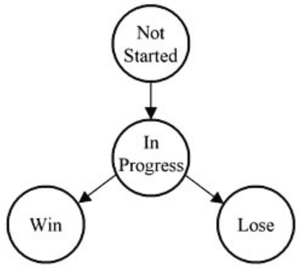


Fig. 2. Automaton of a challenge

of this organization). Consider two challenges a and b and a game where the player must solve a to solve b, a is said to be a *sub-challenge* of b. When the player knows how to solve b, he knows how to solve a. As a consequence any solutions of b includes at least one of the solutions of a.

5.2 Progression of Difficulty

Consider now the type of challenges that are used in games. The choice of challenges and the succession of challenges is related to the flow principle explained in section 2. In many Game Design books, the progression of tension cycles is presented using the Learning/Difficulty curves [30],[31]. At any time of the game, the difficulty of the next challenge must be a little higher than the current level of the player apprenticeship. When he wins the challenge and the tension decreases, the player gets new skills and ability. This correlated progression of skills abilities and difficulty must be kept all along the game.

The same idea is expressed by Jesper Juul using the notion of repertoire of methods [32]

> *'At every instant within a game, a player has created for himself a collection of methods and strategic rules which he has devised and which he applies (the player's repertoire). One strength of a good game is to constantly challenge the player, which in turn leads him to constantly find new strategies, apart of those already in the repertoire'*

There are two ways to control the difficulty: the progression of skills and the mix of challenges. The progression of skills relates the difficulty of a given challenge according to a set of parameters. This notion of difficulty is related to an idea of complexity: what type of problem a human "processor" is able to face taking into account his level of practice. How far can he move buttons, memorize configuration, how precise can be his shot, how long can he stay on one foot? As in any sport or mental exercise, the player's practice enhances his skills, and the same challenge can be solved using parameters chosen to increase the level of difficulty.

The second way to control the difficulty is to mix challenges. The solution of many game challenges relies on mastering of a set of basic techniques and then

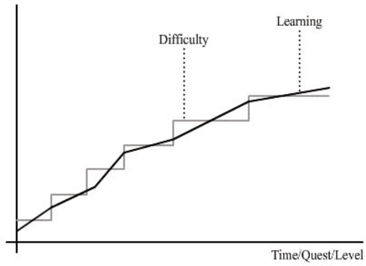


Fig. 3. Difficulty and learning curves

to try to combine them. In strategy games, you master first the strength and movements of units, then the position of production units, then the technological evolution. At each level of a game, the player understands new mechanisms, then slowly combines them. It is the same with basic attacks and combo in fighting games, with group strategy in FPS, and, at last but not least, the increasing complexity of puzzles in adventure games.

Consider the three following challenges: (A) Send a ball in a basket. For a given ball the difficulty of A decrease with the size X of the basket. (B) Press a button when a given event occurs. The difficulty decreases with the accepted error E between the date of the event and the moment when the button is pressed. (C) For given X and E, sending a ball in a basket when a given event occurs is more difficult than A followed by B.

We may state that in a game, the progression of difficulty relies on two sets of challenges:

- A set of basic challenges whose complexity, can be controlled through a set of parameters.
- An ordered set of challenges. The atoms challenges are of the first type. The solutions of higher level challenges can be deduced from those of lower level challenges.

5.3 The Difficulty Evaluation Function

Let us set up some properties that must fulfill the difficulty of a challenge a, $D(a)$:

- D must be measurable using for example a tool able to record the internal states of the game.
- D must allow comparing the difficulty of two challenges, at least of the same "kind" (jumping in platform game, for example)
- D must be relative to the game history, in particular to the progression of the player's skill according to the set of challenge already overcome.
- D must depend on the time used to finish the challenge.

Let A be the set of all challenges that have been solved before time 0. We define $LOSE(a,t)$ and $WIN(a,t)$ as the following events:

– $LOSE(a, t)$ = the automaton of a reaches the state LOSE before time t, starting at time 0.
– $WIN(a, t)$= the automaton of a reaches the state WIN before time t, starting at time 0.

We propose a definition of the difficulty D as a conditional probability:

$$D(a, t) = Probability\{LOSE(a, t)/A\}$$

The Easiness E of a can be also defined in the same way:

$$E(a, t) = Probability\{WIN(a, t)/A\}$$

At all time $D(a, t) + E(a, t) \leq 1$. We can also consider the steady state difficulty and easiness:

$$D^*(a) = \lim_{t \to \infty} D(a, t) \text{ and } E^*(a) = \lim_{t \to \infty} E(a, t)$$

If challenge a must necessarily be finished in the game $D^*(a) + E^*(a) = 1$. These functions gives us two kind of information about the challenge difficulty. First, E^* gives us the difficulty of the challenge in term of the probability that a player overcomes it. But we also can be more precise and with $E(a, t)$, get the probability that a player has to overcome the challenge before time t. We assume that designers are able to implement in the game code some triggers associated to the transitions in each challenge automaton during a test performed by players. The time needed to perform a challenge and the fact that a challenge has been successful or not can be recorded.

But in an open game, there is a small chance that two players will reach the same challenge following the same history. Hence the A of the two players will be different. So, it is necessary to drive the tester with a walk-through. In this case the Difficulty and the Easiness can be statistically estimated, and, under some ergodic assumptions $D^*(a)$ and $E^*(a)$ also.

This can lead to validate experimentally numerous assumptions about the learning and the difficulty curves. For example, if a is a sub-challenge of b then $D^*(a) < D^*(b)$. In the same case, if the player plays twice a, even if he loses the first execution, the second one should be less difficult. Lets denote (a *knowing* a) this second challenge:

$$D^*(a \textit{ knowing } a) \leq D^*(a)$$

If a is a sub-challenge of b and if the player has already played a before b then

$$D^*(b \textit{ knowing } a) \leq D^*(b)$$

More generally, this can lead to validate experimentally the theory of learning and difficulty curves.

6 Conclusion

One of the core component of a good gameplay is the good tuning of the challenge difficulty. In this paper, we have presented the link between challenges and player's enjoyment, in term of tension-resolution cycles. However, there is a lack of a general definition of the difficulty in games, methodology and tools to measure it. We have reported a first experiment using a synthetic player. This experiment shows that with a basic AI driven player, we can extract objective difficulty measures out of a simple game. But as we stated before, the player ability to overcome a challenge depends on his experience of the game. Thus, we propose a definition of difficulty taking into account the past experience of the player. This definition relies on the main properties presented in this paper. This function is measurable as long as the game design provides a clear specification of challenges. The next step of our research is to implement, in different types of games, and to experiment the evaluation of the function using real players. This will lead to an experimental validation of the apprenticeship and learning curves.

References

1. Boutros, D.: Difficulty is difficult: Designing for hard modes in games. Gamasutra (2008), http://www.gamasutra.com/ (last access 01/2009)
2. Adams, E.: The designer's notebook: Difficulty modes and dynamic difficulty adjustment. Gamasutra (2008), http://www.gamasutra.com/ (last access 01/2009)
3. Juul, J.: The game, the player, the world: Looking for a heart of gameness. In: Copier, M., Raessens, J. (eds.) Level Up: Digital Games Research Conference Proceedings, pp. 30–45 (2003)
4. Hunicke, R.: The case for dynamic difficulty adjustment in games. In: Advances in Computer Entertainment Technology, pp. 429–433 (2005)
5. Eco, U.: L'oeuvre ouverte. Seuil (1965)
6. Koster, R.: A Theory of Fun for Game Design. Paraglyph Press, Scottsdale (2005)
7. Sweetser, P., Wyeth, P.: Gameflow: a model for evaluating player enjoyment in games. Comput. Entertain. 3(3), 3 (2005)
8. Yannakakis, G.N., Hallam, J.: Towards optimizing entertainment in computer games. Applied Artificial Intelligence 21(10), 933–971 (2007)
9. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper Perennial (March 1991)
10. Jsselsteijn, W., de Kort, Y., Poels, K., Jurgelionis, A., Belotti, F.: Characterising and measuring user experiences in digital games. In: International Conference on Advances in Computer Entertainment Technology (2007)
11. Andrade, G., Ramalho, G., Santana, H., Corruble, V.: Extending reinforcement learning to provide dynamic game balancing. In: IJCAI 2005 Workshop on Reasoning, Representation, and Learning in Computer Games, pp. 7–12 (2005)
12. Scott, B.: Architecting a game ai. In: AI Game Programming Wisdom 1, Charles River Media, Inc. (2002)
13. Macleod, A.: Game design through self-play experiments. In: ACE 2005: Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology, pp. 421–428. ACM, New York (2005)
14. Kirby, N.: Ai as gameplay analysis tool. In: Game Programming Wisdom 4. Course Technology, Cengage Learning, pp. 39–49 (2008)
15. Bullen, T., Katchabaw, M.J., Dyer-Witheford, N.: Automating content analysis of video games. In: Canadian Game Studies Association (CGSA) Symposium (2006)

16. van Lankveld, G., Spronck, P., Rauterberg, M.: Difficulty scaling through incongruity. In: Proceedings of the Fourth Artificial Intelligence and Interactive Digital Entertainment Conference, Stanford, California, USA, October 22-24 (2008)
17. Nantes, A., Brown, R., Maire, F.: A framework for the semi-automatic testing of video games. In: Artificial Intelligence and Interactive Digital Entertainment Conference. AAAI, Menlo Park (2008)
18. Chan, B., Denzinger, J., Gates, D., Loose, K., Buchanan, J.: Evolutionary behavior testing of commercial computer games. In: Congress on Evolutionary Computation, CEC 2004, June 2004, vol. 1, pp. 125–132 (2004)
19. Spronck, P.: Evolving improved opponent intelligence. In: GAME-ON 3rd International Conference on Intelligent Games and Simulation, pp. 94–98 (2002)
20. Laird, J.E.: Game developers magazine (2000)
21. Pittman, J.: The pac-man dossier. Gamasutra (2009), http://www.gamasutra.com/ (last access 01/2009)
22. Yannakakis, G.N., Hallam, J.: Evolving opponents for interesting interactive computer games. In: Animals to Animats 8: Proceedings of the 8th International Conference on Simulation of Adaptive Behavior (SAB 2004), Santa Monica, CA, USA, pp. 499–508. MIT Press, Cambridge (2004)
23. Lucas, S.M.: Evolving a neural network location evaluator to play ms. pac-man. In: Proceedings of the 2005 IEEE Symposium on Computational Intelligence and Games, CIG 2005 (2005)
24. Gallagher, M., Ledwich, M.: Evolving pac-man players: Can we learn from raw input? In: Computational Intelligence and Games, CIG (2007)
25. Bonet, J., Stauffer, C.: Learning to play pac-man using incremental reinforcement learning (2001), http://www.ai.mit.edu/-people/-stauffer/-Projects/-PacMan (accessed December 5, 2008)
26. Rosca, J.P.: Generality versus size in genetic programming. In: Genetic Programming 1996: Proceedings of the First Annual Conference, pp. 381–387. MIT Press, Cambridge (1996)
27. Gallagher, M., Ryan, A.: Learning to play pac-man: an evolutionary, rule-based approach. In: Evolutionary Computation (CEC 2003), vol. 4, pp. 2462–2469 (2003)
28. Watkins, C.J.: Learning from Delayed Rewards. PhD thesis, Cambridge (1989)
29. Natkin, S., Vega, L.: A petri net model for computer games analysis. Int. J. Intell. Games & Simulation 3(1), 37–44 (2004)
30. Natkin, S., Delocque-Fourcaud, A.M., Novak, E.: Video Games and Interactive Media: A Glimpse at New Digital Entertainment. AK Peters Ltd. (2006)
31. Byrne, E.: Game Level Design. Game Development Series. Charles River Media (December 2004)
32. Juul, J.: Half-Real: Video Games between Real Rules and Fictional Worlds. MIT Press, Cambridge (2005)

Dance Motion Control of a Humanoid Robot Based on Real-Time Tempo Tracking from Musical Audio Signals

Naoto Nakahara[1], Koji Miyazaki[1], Hajime Sakamoto[1], Takashi X. Fujisawa[1], Noriko Nagata[1], and Ryohei Nakatsu[2]

[1] Kwansei Gakuin University, School of Science and Technology
2-1 Gakuen, Sanda, 669-1337 Japan
Nakahara-N@kwansei.ac.jp
[2] National University of Singapore
21 Lower Kent Ridge Road, 119077 Singapore
idmdir@nus.edu.sg

Abstract. This paper proposes a system that controls and generates a humanoid robot's dance motion in real-time using the timing of beats in musical audio signals. The system tracks changes in tempo and calculate the integration value of a decibel by analyzing audio signals in real-time. It uses the information to add changes to the robot's dance motion. Beat intervals and the integration value of decibels are used to change the tempo and range of the robot's dance motion respectively. We propose a method to synchronize dance motion of robot with musical beat, changing the robot's dance motion interactively according to the input value.

Keywords: Robot, Dance Motion, Beat Tracking, Music Understanding, Human Computer Interaction.

1 Introduction

Music and dance have had a strong relationship since ancient times. People started to move their bodies as music was played. Dance was used as a way to express feelings and communicate with each other.

There is numerous research on retrieving information from musical audio signals, but few applied the information to contents such as humanoid robots. There is some research that applies musical information to robot motion, but those contents dealt with comparatively slow movements and the information was presented one-sided from robot to human. We think that a bi-directional information exchange is an important factor for human-robot interactions. It is important to synchronize robotic motion with music, utilizing faster motion speed, to be able to perform more complex movements.

We implemented a system that tracks the timing of beats from audio signals that are input by either .wav file format music files or a keyboard. The system can change and control tempo and range of robotic dance motion by using the information extracted from audio signals in real-time. As the music is played by the user, the robot

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 36–47, 2009.

synchronizes its dance motion with music's tempo and beat. The user can control the tempo and range of motion of the robot by tapping keys at different speeds and strength.

2 Related Work

Research done by Goto [1], [2] is famous for implementing a real-time beat tracking system called "BTS." BTS can track the timing of beats from musical audio signals. This model utilizes a multi-agent structure, where each agent predicts the time period between beats (inter beat interval) and the time of next beat with different parameters. Since this model tracks the timing of beats in music that have a roughly constant tempo, it takes a long time or simply can not track tempo changes. Research done by Yoshii [3] proposes a method to synchronize a humanoid robot's (ASIMO [4]) steps with musical beats. The musical signal is inputted from a microphone on ASIMO's ear and extracts beat intervals from that signal. Then ASIMO tries to synchronize his steps to that music. The step interval is limited to between 1000 and 2000 ms, and can not synchronize fast moves to the music. The movement is limited to the basic movement of a human, which is a stamping motion and does not change. Shiratori researched a method that allows a dancing robot the ability to observe and imitate human dance performances, making the movement more natural [5]. They extracted a sequence of primitive motion acquired from motion data. A method to convert captured human motion to a feasible robot motion automatically was proposed. In addition, they proposed a method to generate new combinations of motion by extracting the timing of beats and the dynamics of music along the time axis and the mapping sequence of primitive motion. These calculations are done offline, not real-time.

Although research pertaining to synchronizing motion of humanoid robots and music has been done, the motion of the robot is still limited to slow and basic motions of humans, or does not work in real-time. An existent real-time beat tracking model from audio signals takes a long time or can not track tempo changes. In addition, input is only received from a music file where information is presented one-sided, from robot to human. To realize an interactive robot dancer, the system must synchronize the robot's comparatively fast dance motion to musical audio signals and change motions according to the input signal in real-time. To track tempo changes in music faster than existing real-time beat tracking models are required when synchronizing robotic dance motion to audio signals that are inputted by a user, for example from a keyboard.

We propose a method to synchronize robotic dance motion with musical audio signals and change the robot's motion interactively in real-time according to the input audio signal. A method to track tempo changes faster than a past model is also proposed.

3 Retrieving Tempo and Beat from Musical Audio Signal

At the musical audio signal analysis stage of our system, we implemented a beat-tracking model to track the basic timing of beats in music audio signals. Although beat tracking models proposed so far dealt with songs that have a roughly constant tempo, we implemented an algorithm which can track tempo changes in audio signals faster than the existent model. The system runs on a cluster of a duel processor, 2.16GHz operation machine.

In the frequency analysis stage, musical audio signals are continuously transferred into spectrograms by applying FFT. In our implementation, the FFT size is 1024 samples, the shifting interval is 256 samples, and input signals are sampled at 22050Hz. Therefore, the spectrogram is obtained every 11.6msec, and frequency resolution is 21.53Hz.

3.1 Finding Onset-Components

An important clue to finding the beat timing is based on the general knowledge of music; that at most times the instruments are sounded at the timing of the beat. This is the basic view of the beat tracking algorithm. Algorithms based on this make different predictions to output the best prediction of the next beat interval. If the Formula.1 is fulfilled, Formula.3 is used to calculate the power of onsets on a specific frequency. If Formula.1 is not fulfilled, the power of the onset component is considered as zero. The value of *prevPower* is given by Formula.2. Fig.1 shows an image of an onset component. An onset component is a frequency component that is likely derived from an onset. $p(t, f)$is the power at time t and frequency bin f, and $d(t, f)$ is a power of the onset component at t and f. Unit of time t is 11.6msec, which is time resolution of FFT and the unit of frequency f is 21.53Hz which is the frequency resolution of FFT in the present implement.

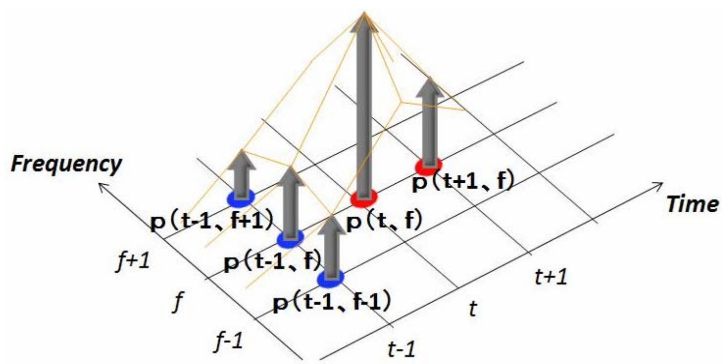


Fig. 1. Extracting onset component

$$\min(p(t,f), p(t+1,f)) > prevPower \quad (1)$$

$$prevPower = \max(p(t-1,f), p(t-1,f \pm 1)) \quad (2)$$

$$d(t,f) = \max(p(t,f), p(t+1,f)) - prevPower \quad (3)$$

Onset components powers are classified into seven frequency ranges (0-125Hz, 125-250Hz, 250-500Hz, 0.5-1kHz, 1-2kHz, 2-4kHz, and 4-11kHz). All of the components powers are summed according to the classified range, described as seven onset component powers. These seven powers are recorded every 11.6msec along the time axis and described as the onset component's power vector. Each power vector is smoothed by the Savitzgy-Golay smoothing algorithm along the time axis **[6]**. The onset time is given by its time that gives the local peak of the onset power vector, and they are described as onset time vectors. Furthermore, seven onset time vectors are transferred into three types of vectors according to frequency focus type, which is low, middle, and all.

3.2 Auto-Correlation and Cross-Correlation

As a next step, onset time vectors are used as clues to predict IBI(inter beat interval), and the next beat time. Fig.2 shows an image of predicting IBI and next beat time. IBI is the inner beat interval which is the time between two beats. The IBI of the audio signal is obtained by calculating auto-correlation of onset time vectors. Next, beat times are predicted by calculating cross-correlation using onset time vectors. The system uses different parameters when calculating auto-correlation and cross-correlation to get different interpretations. Calculations of auto-correlation and cross-correlation on several different parameters and the calculation of interpretation are implemented as different agents in the system. Each agent has different parameters, such as auto-correlation window size and onset time vector frequency focus type. When the auto correlation window size becomes small, the agent's sensitivity to temporary changes will improve, but stability of the prediction will be low.

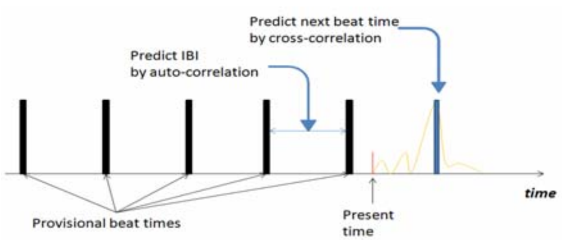


Fig. 2. Prediction of inter-beat interval and the next beat time by calculating auto-correlation and cross-correlation using onset vectors and provisional beat times

3.3 Integration of Agents

The difficult part of tracking tempo and beat is that all instruments are not always sounded at the timing of the beat and that appropriate parameters for calculating auto-correlation differs by music. To cope with the ambiguity of music, the system uses multi-agent predictions and evaluations. Each agent has functions that evaluate their own prediction. All agents are grouped according to their predicted next beat time. Then, evaluation values of each agent are summed in the group to obtain each group's evaluation value. The group that has the highest evaluation value is selected as the primary group. The agent that has the highest evaluation value in that group is considered the most reliable agent. The system outputs the prediction of the most reliable agent as a final output.

4 Robotic Motion Control

At a robotic motion control stage of our system, robotic motion is generated and controlled. Basic functions and commands to control and synchronize with music are sent from a PC program to the robot's micro computer.

4.1 Humanoid Robot "Tai-chi"

Humanoid robots that are used in the system are "Tai-chi." Tai-chi is a humanoid robot that is made by Nirvana Technology. Tai-chi's height is 37cm and weight is 2.2kg. Specifications of Tai-chi are shown in Table.1

Fig. 3. Humanoid robot "Tai-chi"

Table 1. Specification of Tai-chi

Size/Weight	37cm/2.2kg
Degree of flexibility	21 (12/legs, 8/arms, 1/head)
CPU	SH2/7046 50MHz
Motor	KRS-2346ICS PDS-947FET
Battery	NiCad battery RCP-33 7.2V 1100mAh

4.2 Dance Key Pose Database

To preserve the basic dance motions of the robot, the system maintains key poses of dance motions. A robotic motion editor is used to produce key poses for the robot.

Image of the motion editor is shown in Fig.4. Tai-chi has 21 joint motors and the motion editor is used to preserve and control each angle of the motor which represents one key pose of the robot's dance motion. Users can easily change the angle of the motors with scrolling bars on the motion editor screen. Also, users can check what kind of key pose is generated with those motor angles by sending motor angles to the robot, or simply watching the computer graphics of the robot on the screen to simulate the robot's key poses and dance motion. The robot' dance motion key poses are described as 21motor angles; these values are saved as text files. Those text files are maintained in the system in the dance motion key pose database.

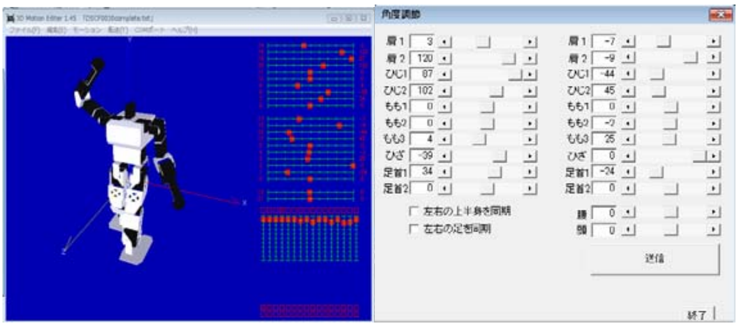


Fig. 4. Motion Editor of Tai-chi

4.3 Producing Dance Motion

An image of produced dance motions is shown in Fig.5. Dance motion is produced by using motor angles of key poses that are loaded from text files in the database. First, 21 motor angles that describe key pose are sent from the PC to the micro computer in the robot, the robot then takes that pose. Next, another 21 motor angles and specific times are sent to the micro computer. This time is the time between one key pose and another in msec. When these values are sent to the micro computer on the robot, it calculates and interpolates each angle of the motors to take the next pose in that specific time from the present key pose. By continuing this process, the robot's dance motion is generated and played. Also, by changing combinations of key poses and selecting different key poses every time, the system can generate many types of dance movements.

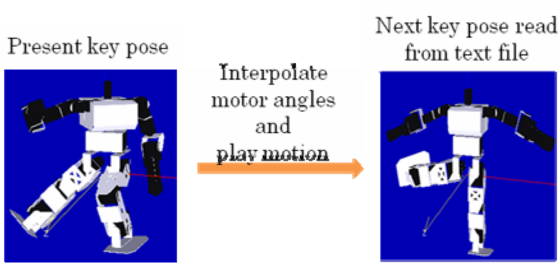


Fig. 5. Playing dance motion

5 Synchronizing Robotic Dance Motion with Music

An algorithm to synchronize robot's dance motion with music is shown in this section. This is realized by sending information from the musical audio analysis stage to the robotic dance motion control stage. The timing of a beat is known as the most fundamental and important factor of the music; it is also important when synchronizing dance motion with it.

5.1 Connection between the Musical Audio Signal Analysis Stage and the Robotic Motion Control Stage

In our present system, the musical audio analysis stage is implemented with Max-MSP; robot dance motion control stage is implemented with Visual C++. Both stages are in the same computer and are connected with TCP/IP. Control signals from the robot dance motion control stage in PC are sent to the robot through USB. Although in our present system both stages are in the same computer, it can be separated to different computers. For example, the music audio analysis stage in one computer can be connected with the robotic dance motion control stage in another computer using a standard TCP/IP network.

5.2 How to Synchronize Robotic Motion with Music

To be able to synchronize with the beat, the system should be able to determine the beat before it arrives. The timing of the next beat is predicted from audio signals at the musical audio signal analysis stage and it is transferred to robot dance motion control stage.

How to synchronize robotic motion with music is shown in Fig.6. When to send information and what kind of information to send is important in synchronizing robotic dance motion with music. When to send information from the musical audio analysis to the robotic motion control stage is decided before the next beat time comes. In the musical audio analysis stage, time between beats and the next beat time is predicted and calculated as final output. When the next predicted beat time comes, IBI (time period between adjacent beats) is sent from the musical audio analysis to robotic motion control stage. When the robotic motion control stage gets IBI from the present music, it selects the next key pose and reads the 21 motor angles from a text file. Also, the target time for the robot to transfer to next key pose from the present key pose is calculated. Target time is calculated by adding the IBI to the time when the information is transferred to the robotic motion control stage. Then, the 21 motor angles and the target time for the robot to transfer to the next pose is sent to the micro computer on the robot through a USB line. When the micro computer receive the information, it sets up the target time and next key pose motor angles, and calculates and interpolates each angle of motors as the time passes to transfer to that key pose at target time from the present pose. The robot's dance motion is generated and played by continuing this process.

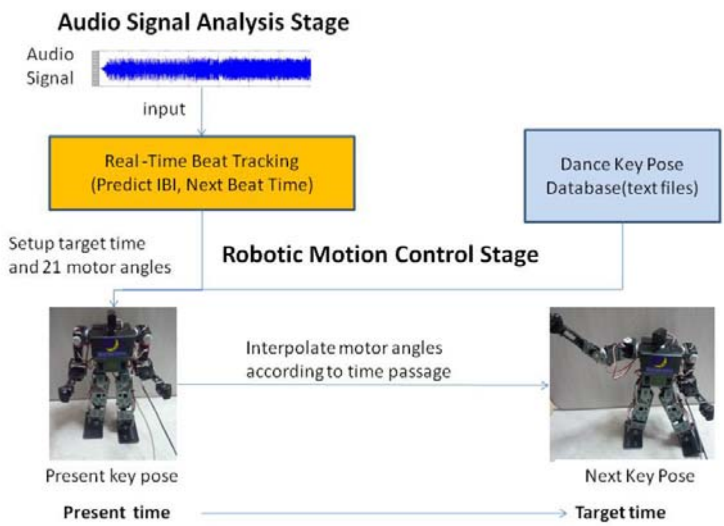


Fig. 6. Synchronizing robot's motion with music

6 Interaction with Robot Using Keyboard

The system works in two modes which are the "appreciation mode" and the "interactive mode." In the appreciation mode, .wav file format music files are inputted and the music has roughly constant tempo. Since the input signal has a roughly constant tempo, the system does not have to consider tempo changes. In the "interactive mode," audio signals that are obtained by the user's input from the MIDI keyboard are analyzed as input. Since the user might change the speed of key tapping, the system has to consider tempo changes and synchronize the robot's dance motion. Users can select the timbre of instruments such as piano and drums and play some chords on the keyboard or a tap of a key in a certain tempo. Although the system uses a MIDI keyboard as an input device, it does not use any MIDI information. MIDI signal is transferred to raw audio signal by connecting output jack and input jack on a PC with a line, and the signal from the input jack is analyzed as input of the interactive mode. The integration value of the decibel value of input signal is used to decide the strength of key tapping. Robotic motion is selected and changed according to the tempo and integration value of the decibel of the audio signal.

6.1 Tracking Musical Tempo Changes

As explained in section 3.2, the calculation of the auto-correlation and cross-correlation is used to predict the IBI of the music and next beat time. If the input is the musical audio signal from a music file, the tempo is roughly constant and does not have to consider tempo changes, so is important to maintain the prediction of the IBI as stably as possible. To make the stability higher, the system produces a history of IBI that are predicted by calculating the auto-correlation of onset time vectors.

Auto-correlation is calculated every 11.6msec and the system maintains every prediction of IBI as a history of the signal analysis. Using the history of the IBI, the most frequent value of the IBI is predicted by calculating the auto-correlation output from the musical audio analysis stage. The system uses IBI history to prevent the output of temporally wrong predictions of IBI and lower the sensitivity of changes in prediction by calculating auto-correlation. In the interactive mode, the input is an audio signal that is obtained from user's keyboard which allows the change of tempo to be assumed. To track the tempo changes in audio signal, the system must be sensitive to changes, so another algorithm is needed. If the input is from the user's keyboard, the audio signal analysis stage must maintain stability in predicting the IBI and be sensible to tempo changes. To be sensible to tempo changes, the system resets the history of the IBI if the IBI prediction is calculated and auto-correlation fulfills the following two conditions. One is that the predicted IBI differs by ±50 msec compared with the frequent IBI in the history. Another is that the predicted IBI is within ±50 msec compared with the previous predicted IBI. The system decides that the tempo of the musical audio signal has changed if these two conditions are fulfilled five times in a row. By using this algorithm, the system can maintain stability to temporal tempo changes and decrease incorrect predictions as well as track main tempo changes in the audio signal.

6.2 Motion Change of the Robot by the Integration Value of Decibels

In an interactive mode, audio signals from the keyboard are used as input. Users can control the tempo of the robot's dance motion by changing the tempo of tapping the key. Also, the user can control the range of dance motion by changing the strength of key tapping. This is realized by calculating the integration value of the decibel of audio signals from the keyboard and selecting key poses that correspond to that value. Motor angles of four key poses are stored in one text file as one unit. Balances of the robot through transitions of key poses in units are checked by the motion editor. Last key pose in an unit is set as a standing pose, which is a neutral pose of the robot, so that the robot can maintain the balance through the transition to another motion unit. Each dance motion text file in the database is classified into three groups according to the range of motion. In the present implementation, there are five units of motion for each group, which are fifteen units and sixty key poses in total. The group the motion text file belongs to is decided by the range of value changes of each motor angle. When the integration value of the decibel is obtained from the audio signal, the system determines which group of key pose text files to select from. The value is rescaled between 0 and 1. Possibility of appearance of each group is determined by sending the value to fuzzy functions and one group is selected using that possibility. Fuzzy functions are used to emulate the ambiguity of humans when selecting dance motion. The system finally selects the next text file of dance motion to play from that group. If the system selects the same group as last time, next text file is selected randomly in the same group. By continuing this process the user can control the range of the robot's dance motion using a keyboard and changing the strength of key tapping.

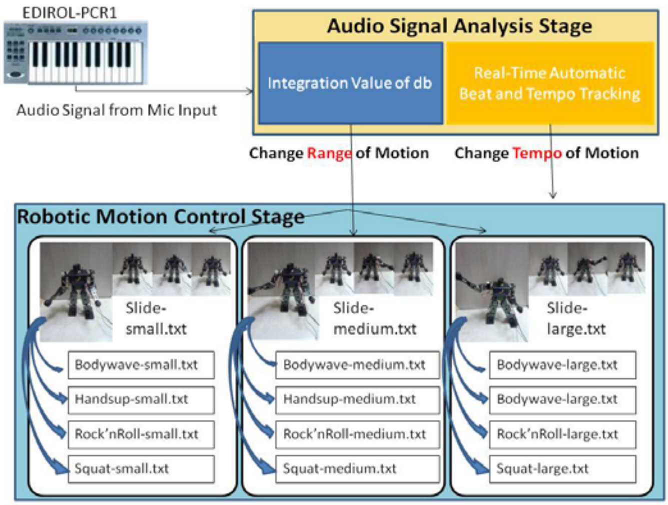


Fig. 7. Changing motions of the robot by the integration value of decibel and tempo of audio signal

7 Evaluation

7.1 Beat Tracking Result

Ten songs were used to evaluate audio signal analysis part of our system. We retrieved the time of beat in music by using the software called "AudaCity"**[7],** which can play and display the waveform of the music at the same time. Then we compared the time of each beat with the time of beat retrieved from beat tracking part of the system. Thirty seconds, from twenty seconds to fifty seconds of each song were used in the evaluation. Table2 is a list of ten different songs, and an average error of every beat in a song. Average of all ten song's average errors were 36.1msec.

Table 2. Result of the Beat Tracking

Title	Artist	Type	BPM	Average Error
Love me do	The Beatles	Rock	147	37msec
Rock 'N' Roll Star	Oasis	Rock	139	32msec
Holiday	Green Day	Rock	147	26msec
Movin' on without you	Hikaru Utada	Pop	122	39msec
Together	EXILE	Pop	112	54msec
Another Story	Mr.Children	Pop	91	36msec
FAKE	Mr.Children	Pop	125	32msec
Dream Fighter	Perfume	Techno-pop	135	19msec
BANZAI	MISA	Trance	145	42msec
Nobody Knows	DARK-ONE	Trance	145	45msec

7.2 Dance Motion Made by the System

First we used a music file as an input. Without using the motion change method, the robot just danced to the beat of the music in a single motion, but by using the algorithm proposed above, various combinations of motions were played according to the change of the input signal, instead of just playing the same motion many times. The motion change algorithm helped to reduce the monotony of dance motions.

Next, we used the keyboard as an input device and played some chords on it. The robot started to dance on the tempo of the tapping of the keyboard. If the keyboard is tapped weakly, the dance motion of the robot changed to small motion, and if tapped strongly, the dance motion changed to large motion. If we changed the tapping of the keyboard slower to faster or faster to slower, the robot tracked and changed the tempo of the dance in 5 to 10seconds.

8 Conclusion

We developed a system that controls and generates a humanoid robot's dance motion in real-time using the timing of beats in musical audio signals. The beat is extracted from the musical audio signal in real-time, allowing the measurement of intervals between beats and the prediction of the next beat. By taking history of IBI, and clearing the history if certain conditions are fulfilled, the system maintains stability of prediction and tracks tempo changes faster than past real-time beat tracking models. We also proposed an algorithm to change the tempo and the range of dance motion of a robot interactively by using keyboard as an input device and changing the tempo and strength by key tapping. By using this algorithm, the system can interactively change the dance motion of robot according to the input and reduce the monotony of generated motions.

References

1. Masataka, G., Yoichi, M.: Real-time Beat Tracking for Drumless Audio Signals –Chord Change Detection for Musical Decisions, Speech Communication, 311–335 (1999)
2. Masataka, G.: An audio-based real-time beat tracking system for music with or without drum-sounds. Journal of New Music Research 30(2), 159–171 (2001)
3. Takaaki, S., Atsushi, N., Katsushi, I.: Detecting Dance Motion Structure through Music Analysis. In: IEEE International Conference on Automatic Face and Gesture Recognition, pp. 857–862 (2004)
4. An Official Site of ASIMO, `http://www.honda.co.jp/ASIMO/`
5. Kazuyoshi, Y., Kazuhiro, N., Toyotaka, T., Yuji, H., Hiroshi, T., Kazunori, K., Tetsuya, O., Hiroshi, O.: A Biped Robot that Keeps Steps in Time with Musical Beats while Listening to Music with Its Own Ears. In: International Conference on Intelligent Robots and Systems, pp. 1743–1750 (2007)
6. Abraham, S., Marcel, J.E.G.: Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry 36(8), 1627–1639 (1964)
7. An Official Site of Audacity, `http://audacity.sourceforge.net/`

8. Shinichiro, N., Atsushi, N., Kazuhito, Y., Hirohisa, H., Katsuhi, I.: Generating Whole Body Motions for a Biped Humanoid Robot from Captured Human Dances. In: IEEE International Conference on Robotics and Automation (2003)
9. Dixon, S.: A Beat Tracking System for Audio Signals. In: Proc. of Diderot Forum on Mathematics and Music, Vienna, Austria (1999)
10. Simon, D.: A Lightweight Multi-Agent Musical Beat Tracking System. In: AAAI Workshop on Artificial Intelligence and Music (2000)
11. Shinozaki, K., Oda, Y., Tsuda, S., Nakatsu, R., Iwatani, A.: Study of dance entertainment using robots. In: Pan, Z., Aylett, R.S., Diener, H., Jin, X., Göbel, S., Li, L. (eds.) Edutainment 2006. LNCS, vol. 3942, pp. 473–483. Springer, Heidelberg (2006)
12. Wama, T., Higuchi, M., Sakamoto, H., Nakatsu, R.: Realization of Tai-chi Motion Using a Humanoid Robot. In: Rauterberg, M. (ed.) ICEC 2004. LNCS, vol. 3166, pp. 14–19. Springer, Heidelberg (2004)
13. Foote, J.: Content-Based Retrieval of Music and Audio. In: Multimedia Storage and Archiving Systems II, Proceedings of SPIE, pp. 138–147 (1997)
14. Scheirer, E.D.: Tempo and beat analysis of acoustic musical signals. Journal of Acoust. Soc. Am. 103(1), 588–601 (1997)

Marker-Less Tracking for Multi-layer Authoring in AR Books

Kiyoung Kim, Jonghee Park, and Woontack Woo*

GIST U-VR Lab.
500-712, Gwangju, S. Korea
{kkim,jpark,wwoo}@gist.ac.kr

Abstract. An Augmented Reality (AR) book is an application that applies AR technologies to physical books for providing a new experience to users. In this paper, we propose a new marker-less tracking method for the AR book. The main goal of the tracker is not only to recognize many pages, but also to compute 6 DOF camera pose. As a result, we can augment different virtual contents according to the corresponding page. For this purpose, we use a multi-core programming approach that separates the page recognition module from the tracking module. In the page recognition module, highly distinctive Scale Invariant Features Transform (SIFT) features are used. In the tracking module, a coarse-to-fine approach is exploited for fast frame-to-frame matching. Our tracker provides more than 30 frames per second. In addition to the tracker, we explain multi-layer based data structure for maintaining the AR book. A GUI-based authoring tool is also shown to validate feasibility of the tracker and data structures. The proposed algorithm would be helpful to create various AR applications that require multiple planes tracking.

Keywords: augmented reality, marker-less tracking, layer authoring, page recognition, AR book, SIFT.

1 Introduction

Camera tracking plays an important role in the implementation of Augmented Reality (AR) applications. Generally, the purpose of camera tracking in AR is to compute a relative camera pose, represented in a rotation and a translation matrix, with respect to the local coordinates of a tracked object. Then, the camera pose is used to augment virtual contents with the projection matrix obtained from camera intrinsic parameters. For AR book applications it is mandatory to recognize which page is visible and to compute a camera pose in real time. In particular, AR book applications requires a robust page number extraction because it allows developers to map different virtual contents onto the corresponding page of a book. The stability and accuracy of camera poses are also important issues as well as the speed of tracking. An unstable and slow tracker

* This research was supported by the CTI development project of KOCCA, MCST in S.Korea.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 48–59, 2009.

may cause users to lose immersion to the AR books. Additionally, a real-time page recognition method is also requires for the AR books. With AR books, users can not only read the story written in the traditional way, but also view and manipulate 3D models with the help of a good camera tracker. Moreover, AR books offer improved user experiences by providing new applications which were not available with traditional books, such as 3D virtual games or storytelling.

Many AR books with camera trackers have been developed [1,2,3]. In the early stage, fiducial markers were used to get a camera pose relative to the markers [1]. The fiducial markers are convenient for identifying each page and computing camera poses. Additionally, attaching the markers on the physical tools provides a way of interaction which allows direct manipulation of augmented objects. However, the markers distract users' concentrations and are sensitive to occlusions. Recent AR books adopt tracking by detection-based marker-less tracking methods [2][3]. The marker-less methods are robust to occlusions and do not require any distinctive markers onto pages. However, most of the marker-less tracking methods require high computational costs and are not much faster than marker-based methods. Moreover, the fast method, like [4] requires a lot of time and computer memory in training phase so that it takes long time for authoring the contents. A hybrid method which uses markers for page recognition and randomized trees for camera tracking [4] was proposed in order to support multiple pages [5].

In this paper, we propose a new marker-less tracking method which supports real time page recognition and fast camera tracking. We use a multi-core programming approach which separates a page recognition from a camera tracking module in order to improve the performance without losing accuracy and robustness of the tracker. Our method adopts highly distinctive Scale Invariant Features Transform (SIFT) features [6] for the page recognition. We extract SIFT features from each page and save them as a template in the offline process. When an input image comes into the tracker, the tracker compares the SIFT features of the input image and the saved SIFT features template. The processing time for the comparison process is varied according to the total number of features in each page. We proposed an efficient voting-based method to filter out irrelevant pages to reduce the processing time for the recognition. In a camera tracking phase, we perform frame-to-frame matching based on a coarse-to-fine approach with FAST corners [7]. In addition, we explain a multi-layer, a set of subregions of a page, data structure for maintaining the AR book. The Graphical User Interface (GUI) based authoring tool is shown to validate feasibility of the proposed tracker and the multi-layer based data structure.

2 Related Work and Background

2.1 Mark-Less Tracking Methods

There have been recent developments on marker-less tracking. We categorize the marker-less tracking into two groups according to its prior conditions: The first one is detection-based tracking (tracking by detection) methods [4,8,9,10]; the second one is SLAM-based methods [11,12,13,14]. While the detection methods have been used mainly in pre-defined object tracking, the SLAM-based methods have been used in unknown environments. Thus, for AR book applications, detection-based tracking methods are more appropriate than SLAM-based methods because we already know which page are included in the AR book. The important problem of the detection-based methods is the processing time for the feature recognition. Trackers with SIFT [6] or SURF [15] was not enough for real time AR applications due to heavy computational costs. To provide real time recognition, the randomized tree (RT) was adopted [4,10]. However, the RT generation takes around 1 minute per one page. And it is not easy to utilize RT method in a multiple pages recognition because of the large amount of memory for keeping RT structures. It is not desirable characteristic for AR book designers or developers.

Our method overcomes the limitations in training time and the camera frame rate by adopting a multi-core programming approach. The proposed method differs from [14,16] in which we maintain two different SIFT and FAST features for the page recognition and frame-to-frame matching. We get the benefits of both, SIFT features and FAST corners, so that the tracking is done in less than 10ms. Only orthogonal images of the pages are required to start tracking and the preparation (training) time takes less than one second per page. Thus, the proposed method is efficient to use in AR authoring.

2.2 Background

The pinhole camera model is adopted as the camera model for the proposed tracking algorithm. A 3D point $\mathbf{X} = \begin{pmatrix} X, Y, Z, W \end{pmatrix}^T$ in homogeneous coordinates is projected onto $\mathbf{x} = \begin{pmatrix} u, v, w \end{pmatrix}^T$ using Eq. 1:

$$\mathbf{x} = K \begin{pmatrix} R\, t \end{pmatrix} \mathbf{X} \tag{1}$$

where R and t is a rotation and a translation matrix, respectively. K is a camera intrinsic matrix consisting of focal length, principal point and so on. Let us assume that a page of a book is planar. In this case, the Z-value of 3D point $\mathbf{X}$ becomes zero. Thus, the relation between 3D points $\mathbf{X}$ and its projected point $\mathbf{x}$ is represented as a plane transformation, Homography H as shown in Eq. 2:

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \underbrace{K \begin{pmatrix} R_{3\times 2}\; t_{3\times 1} \end{pmatrix}}_{H_{3\times 3}} \underbrace{\begin{pmatrix} X \\ Y \\ W \end{pmatrix}}_{\mathbf{M}} \tag{2}$$

In many AR book applications, it has been assumed that model points $\mathbf{M}$ is known as a priori with the assumption of $W = 1$ without loss of generality. The model points are usually obtained from an orthogonal image of a page by applying existing feature detectors [6,15]. Thus, we should know the correspondences of model points and their observed points in the input image to compute

H. Then, given K, the camera pose $\begin{pmatrix} R \; t \end{pmatrix}$ is decomposed from H by using a rotation property [17].

The goal of the page recognition is to provide adequate model points $\mathbf{M}$ when we compute H. It is very inefficient to use all of model points of all the pages because a large number of model points require considerable processing time to recognize each page as well as computational memory. The proposed method reduces the recognition time by comparing the input image with the candidate pages which have high probability.

3 Proposed Approach

3.1 Preparation and Overall Procedure

We take an orthogonal image ($\mathcal{O}$) of a page as a template. Then, we extract SIFT features ($\mathbf{S}$) from the orthogonal image. And also, we apply FAST corner ($\mathbf{F}$) detector to get additional points on the page. As a result, we have $\mathcal{O}_i, \mathbf{S}_i, \mathbf{F}_i$ of i_{th} page. In addition, we generate pyramid images ($\mathcal{L}_{i,s}$) from $\mathcal{O}_i$ where s is a pyramid level from 0 to 4. We apply the FAST corner detection for each level and keep the output points of each level for frame-to-frame matching. The process is done in offline. Thus, it does not affect the performance of the tracker.

As shown in Figure 1, the proposed marker-less tracker uses two threads. The main thread where the camera tracking is carried out via a frame-to-frame matching determines an overall frame rate. The other module, the page recognition process, is repeatedly performed in the background thread, which is relatively slower than camera tracking in the main thread. We explain details of two modules in Section 3.2 and Section 3.3, respectively.

3.2 Real-Time Page Recognition

In this section, we explain the method that determines a visible page. We adopt SIFT features for the page recognition. The SIFT is highly distinctive so that it can support multiple pages. In addition, it does not require a lot of time to process when we make tracking data compared to other method, like [4], which is an efficient property for authoring of many pages. The output of the method are page ID (pid) and the Homography (H) induced from matching results between the input image and $\mathcal{O}_{pid}$. The problem is, specifically, to find out parameters by comparing SIFT features of input image (S_{input}) with the saved features in the preparation step (3.1) which satisfy Eq. 3.

$$E^* = \underset{H,pid}{\operatorname{argmin}} \| \mathcal{S}_{input} - HS_{pid} \| \tag{3}$$

The goal of our method is to reduce the processing time when an AR book has a large number of registered pages. The first step to achieve the goal is evaluating all pages at every frame with given an input image. We use a voting scheme in the first step. We compute the defined scores at each page, and then sort the pages in a high score order. Then, the dense matching process is performed step

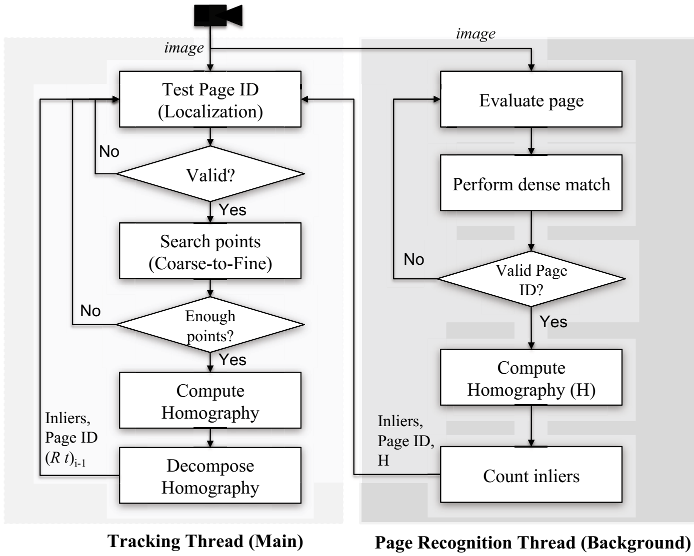


Fig. 1. Overall flowchart of the proposed tracker using two threads

by step from the highest-score-page to the lowest-score-page. The score function of k_{th} page is shown in Eq. 4:

$$\mathbf{Score}(k) = -1 \times \{\omega_1 E_{DIST}(k) + \omega_2 E_{SSD}(k)\} \tag{4}$$

where E_{DIST} and E_{SSD} evaluate a sequential and a similarity constraint, respectively. ω_1 and ω_2 are weighting factors between two evaluation functions. E_{DIST} score measures how much the page is far from the lastly recognized (or currently viewing) page (pid^*) as shown in Eq. 5:

$$E_{DIST}(k) = \|pid^* - k\| \tag{5}$$

Eq. 5 allows to begin with the closest page matching. E_{SSD} score measures how much the input image is similar with the reference pages. We generate low level t pyramid image ($\mathcal{L}_{input,t}$) of the input image. And we perform Sum of Squares Difference (SSD) with $\mathcal{L}_{input,t}$ and each page pyramid image $\mathcal{L}_{k,t}$.

$$E_{SSD}(k) = \mathbf{SSD}\left(\mathcal{L}_{k,t}, \mathcal{L}_{input,t}\right) \tag{6}$$

When the image resolution is 640×480 pixels and $t = 4$, the image size of $\mathcal{L}_{k,4}$ becomes 40×30 pixels. To reduce computational time, we limit the number of candidate pages. If we have the set of page scores $C^* = \{c_i > c_j > c_k, ...\}$, we compute the ratio values between adjacent pages, for example, $r(i,j) = \frac{c_i}{c_j}$. If the ratio is drastically changed, we ignore the consequent pages.

Based on the candidate pages from the process mentioned earlier, we perform dense matching process in the second stage. In the dense matching process, we compare $\mathcal{S}_{input}$ with SIFT features of candidates pages sequentially. We compute Homography and count inliers during the dense matching process. If the number of inliers satisfies our condition, then the system recognize the visible page as the recognized one. After the visible page is identified, we compute Homography and count inliers again. Finally, the page ID, Homography, and inliers information are sent to the main thread. For speeding up matching process, k-d tree can be used. Therefore, we can easily get the page ID without comparing all features in all pages.

3.3 Marker-Less Tracking

In this section, we explain the tracking thread illustrated in Figure 1. The main thread is used for frame-to-frame matching, which matches points from two sequential images. The frame-to-frame matching enables a fast and an easy way to handle points because the movement of points between each frame is short. The tracking thread has i) localization and ii) frame-to-frame matching modules.

The localization is to find the current locations of the saved FAST corners with given Homography (H) obtained from the page recognition process. From the localization process, we find out which points are visible and valid for tracking. This is tested whenever the recognition module passes the results (page ID, H, and inliers) to the main thread. We need to search the corresponding points again because H often yields inaccurate results. For this purpose, we randomly select few sample points and warp patches of them by H. Then, SSD is performed for the selected points. After we find the corresponding points, we perform the guided matching within very narrow area. Then, we accept only if the reprojection error (R_ϵ) and the number of inliers satisfy the below conditions in the validation test.

$$R_\epsilon < T_{init} \ and \ \frac{\#\text{ of Inliers}}{\text{Total }\#\text{ of Points}} > T_{inliers} \tag{7}$$

From the experiments, we found the tracker works well when T_{init} is 2.0 pixels and $T_{inliers}$ is 0.1. The localization enables recovery of the camera pose when tracking fails.

The frame-to-frame matching starts with the Homography (H_{prev}) of a previous frame and matched points. While the localization is performed only when the page recognition results are available, the frame matching is called at every new frame. First of all, we project high-scored 50 points of $\mathbf{F}_{pid}$ by H_{prev} and search the corresponding points to what within a circular area with radius ρ_b. Then, we compute Homography H_c with the coarse matching results. We project many points based on H_c and search again within a reduced circular area with radius $\frac{\rho_b}{2}$. Finally, we obtain the refined Homography $H_{final}$with many matches and warp the patch at each step with given Homographies. To improve the speed of the tracker, only if the initial error is small, we use the matched patches in the current image without warping. The final step is to decompose the H_{final} into rotation and translation matrix [17].

3.4 Multi-layer Authoring

One of obstacles in building AR book applications with a marker-less tracker is positioning virtual contents. In particular, content-aware augmentation, which augments virtual contents in accordance with real figures, texts, or user-defined regions in the page, cannot be realized without analyzing a page layout and layers. In this section, we address a multi-layer authoring scheme to realize a content-aware augmentation efficiently.

We define a page layout $\mathcal{R}_i$ as an arrangement and a style of virtual contents on the page. $\mathcal{R}_i$ of the page has multiple sub-layers $\{r_1, r_2, ..., \}$ which are defined by users. We use a scene graph structure for representing the multiple layers. Users define a sub-layer by dragging or picking a virtual point and allocate contents to the sub-layer. Currently we provide users pictures, text, sound, videos, and 3D CG models as virtual contents. Each layer is defined with several virtual points by the users and the defined layers are saved. Figure 2 shows a scene graph representation of the AR book and its Extensible Markup Language (XML) representation. The AR book consists of multiple pages. Additionally,

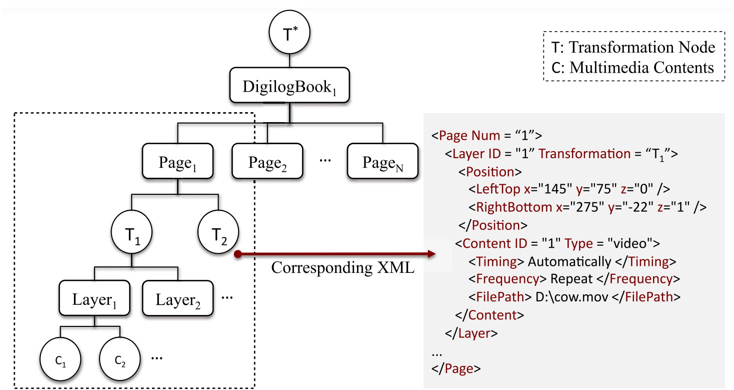


Fig. 2. Illustration of page data structure and its example of XML representation

each page of the AR book is able to include multiple sub-layers which have areas of various contents. Each sub-layer can have multiple contents. When users create a sub-layer or area using an input device, the transformation that includes rotation and translation values is stored in the parent node of the layer.

Figure 3 shows the overall procedure of a multi-layer based authoring. When users make an AR book, a multi-layer authoring system constructs a scene graph of the AR book for rendering virtual contents in an AR environment. The generated multi-layer of the page is stored in XML format to facilitate the modification of the layers later in the layer management module. The AR book Viewer loads XML file according to a visible page ID in online mode. After that, the AR book viewer parses the XML file. After parsing page layers information, contents are augmented on top of the page.

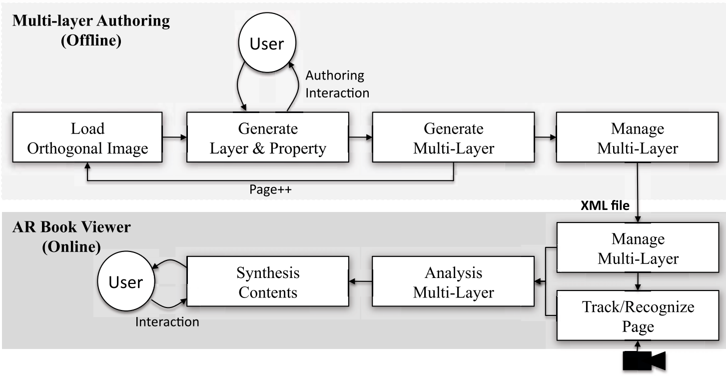


Fig. 3. Procedure for the multi-layer authoring

4 Implementation and Results

In this section, we explain practical implementation issues of the proposed algorithm and show experimental results. We used a 2.66 GHz core 2 duo CPU with a GTX280 NVIDIA graphic card for our experiments. The camera resolution was 640 × 480 pixels and it supported up to 60 frames per second for rendering image onto a screen. We used an ordinary book consisting of figures and texts. No special markers were attached onto pages in the book. We implemented an AR book system using OpenSceneGraph [18]. The OpenSceneGraph allows to manage virtual contents efficiently. And also the proposed tracking algorithms were implemented using OpenCV [19] library. We exploited Graphics Processing Unit (GPU) implementation of SIFT [20] for increasing the speed of the page recognition.

4.1 Page Recognition and Marker-Less Tracker Performance

We measured the processing time in the page recognition. In total 30 pages were registered in advance and the features mentioned in 3.1 were extracted. The time for the registration was only 3 seconds for 30 pages with the help of GPU powers. This is the significant contribution of the proposed tracking algorithm compared to the other tracking algorithm which required about 1 minutes for one page training [4]. Figure 4 shows the selected results when a user turns a page step by step. We augmented different virtual contents with respect to each page ID. The page is successfully detected in all cases. As shown in Figure 4(b), the overall frame rate was more than 30 fps.

The performance results of the page recognition in background thread are shown in Figure 5. We compared the proposed algorithm with the sequential matching test, which showed the worst recognition results. The number of SIFT features depends on the octave and scale parameters. We carried out the comparison with two conditions; light and heavy. As shown in Figure 5, the proposed

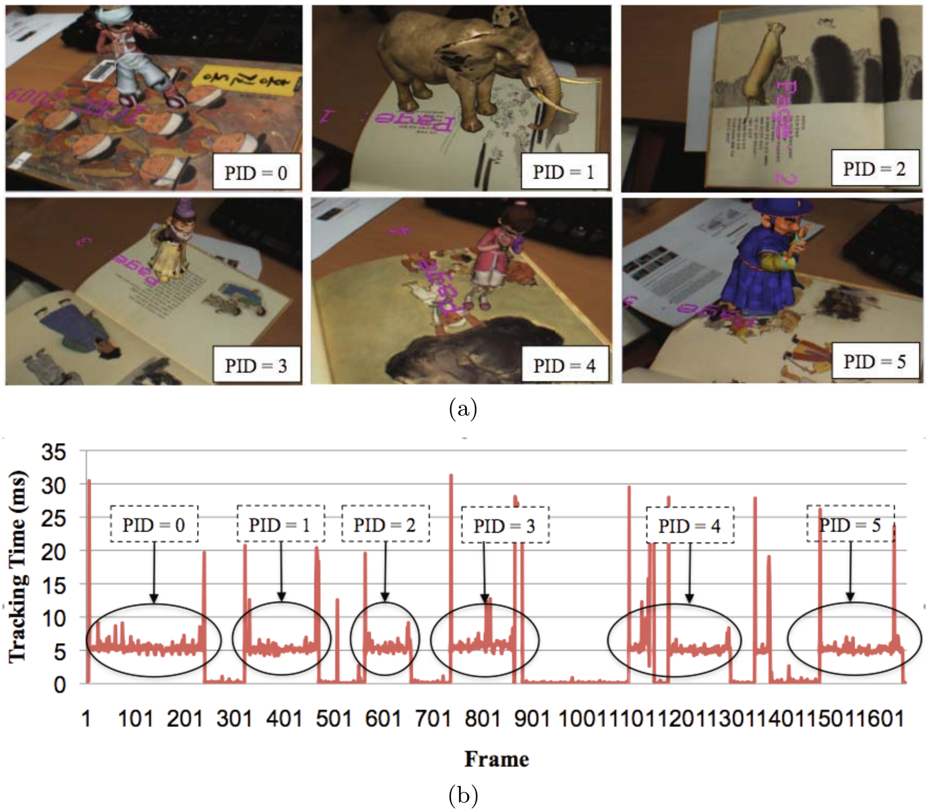
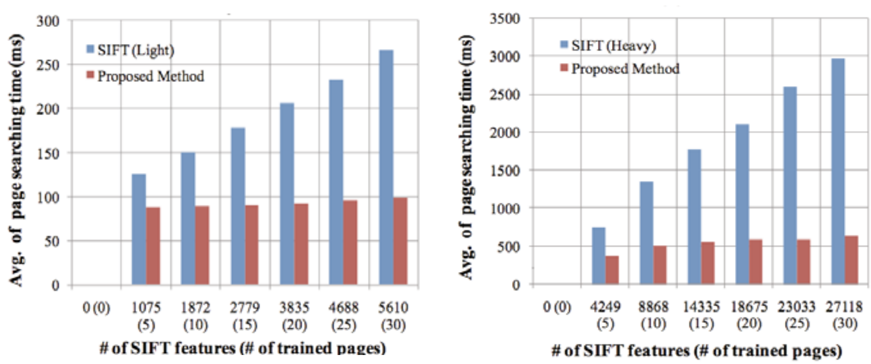


Fig. 4. Results of the page recognition: (a) page-dependent contents and page ID were augmented on each page (b) corresponding performance results for the sequence

method does not drastically increase the page searching time according to the number of SIFT feature points from the registered pages.

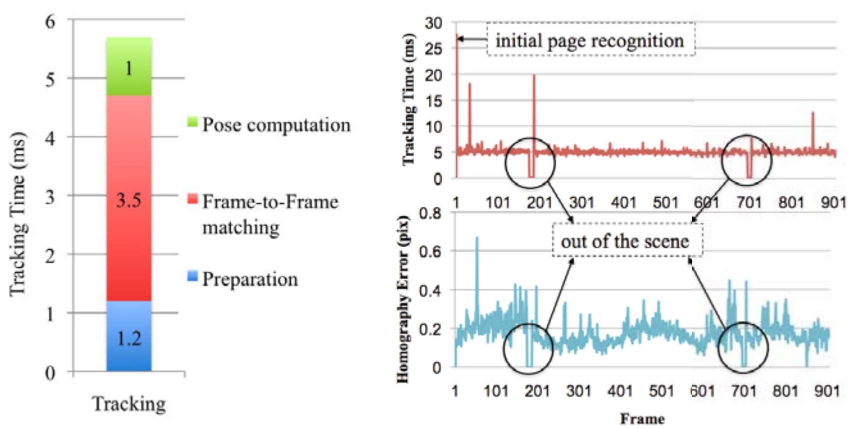
To show the performance of the proposed tracker, we measured the time and the accuracy for the proposed tracker. As shown in Figure 6(a), the tracker is robust to distance changes and rotation of a camera. Figure 6(b) shows the average time spent in each step of the proposed algorithm. The proposed tracker consists of three big steps; preparation, frame-to-frame matching, and pose computation. The preparation step includes capturing camera image, checking the current page number obtained from the background thread, and copying points. Note that the coarse-to-fine matching time depends on the number of tracked points and the size of the searching window. We used the maximum 200 points per page and 8 × 8 window size for this experiment. We observed that our algorithm runs at between 4ms and 8ms in most cases. It is as shown in Figure 6(c).



(a) (b)

Fig. 5. Performance results of the page recognition compared to SIFT sequential matching: (a) SIFT light condition: octave(2) and scale(2) (b) SIFT heavy condition: octave (3) and scale(3)

(a)



(b) (c)

Fig. 6. Results of the marker-less tracker performance: (a) axis augmentation and tracked points (b) average time spent in the camera tracking thread (c) corresponding time and accuracy results

Fig. 7. Snapshots of the multi-layer authoring tool: (from the left) environmental setup, mapping a figure and a video to two layers, example of another page authoring

4.2 Application to Multi-layer Authoring

We built a multi-layer authoring system using the proposed tracker, as explained in Section 3.4. The environmental setup and two examples are shown in Figure 7. We generated two layers and mapped a figure and a video file to each layer. In the authoring window, users could manipulate the defined layers and see the mapped files in a text format. In the AR viewer window, as a result, users could see the AR contents which had been configured before. The benefit of the multi-layer authoring with the proposed marker-less tracking compared to marker-based authoring is that we can use the context of figures and texts on the page when a user make his/her own story.

5 Conclusions and Future Works

We proposed the speed-improved marker-less tracker and the multi-layout authoring method. We showed that the proposed multi-layout authoring was useful to facilitate the content-aware augmentation. Especially, the proposed tracker outperformed the existing methods in time aspect. By using the proposed tracker and authoring tool, we could have a rapid prototype of an AR book application. In the future, we will consider edge features for more robust tracking results. The tracker with a book of more than 30 pages will be tested and analyzed. And also we will consider the interaction authoring with a multi-layer interface. The proposed methods will be used not only in AR book applications, but also in other AR applications which require multiple planes tracking.

References

1. Billinghurst, M., Kato, H., Poupyrev, I.: The magicbook - moving seamlessly between reality and virtuality. IEEE Computer Graphics and Applications 21(3), 6–8 (2001)
2. Taketa, N., Hayashi, K., Kato, H., Noshida, 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)
3. Scherrer, C., Pilet, J., Fua, P., Lepetit, V.: The haunted book. In: IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR 2008, pp. 163–164 (2008)

4. Lepetit, V., Lagger, P., Fua, P.: Randomized trees for real-time keypoint recognition. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, May 2005, vol. 2, pp. 775–781 (2005)
5. Yang, H.S., Cho, K., Soh, J., Jung, J., Lee, J.: Hybrid visual tracking for augmented books. In: Stevens, S.M., Saldamarco, S.J. (eds.) ICEC 2008. LNCS, vol. 5309, pp. 161–166. Springer, Heidelberg (2008)
6. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
7. Rosten, E., Drummond, T.: Fusing points and lines for high performance tracking. In: Tenth IEEE International Conference on Computer Vision, ICCV 2005, September 2005, vol. 2, pp. 508–1515 (2005)
8. Vacchetti, L., Lepetit, V., Fua, P.: Stable real-time 3d tracking using online and offline information. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(10), 1385–1391 (2004)
9. Reitmayr, G., Drummond, T.: Going out: robust model-based tracking for outdoor augmented reality. In: IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR 2006, October 2006, pp. 109–118 (2006)
10. Ozuysal, M., Fua, P., Lepetit, V.: Fast keypoint recognition in ten lines of code. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007, May 2007, pp. 1–8 (2007)
11. Davison, A.J., Reid, I.D., Molton, N.D., Stasse, O.: Monoslam: Real-time single camera slam. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(6), 1052–1067 (2007)
12. Williams, B., Klein, G., Reid, I.: Real-time slam relocalisation. In: IEEE International Conference on Computer Vision, ICCV 2007, September 2007, pp. 1–8 (2007)
13. Castle, R., Gawley, D., Klein, G., Murray, D.: Video-rate recognition and localization for wearable cameras. In: British Machine Vision Conf. (January 2007)
14. Klein, G., Murray, D.: Parallel tracking and mapping for small ar workspaces. In: IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR 2007, October 2007, pp. 225–234 (2007)
15. Bay, H., Ess, A., Tuytelaars, T., Vangool, L.: Speeded-up robust features (surf). Computer Vision and Image Understanding 110(3), 346–359 (2008)
16. Lee, T., Hollerer, T.: Hybrid feature tracking and user interaction for markerless augmented reality. In: IEEE Virtual Reality, VR 2008, February 2008, pp. 145–152 (2008)
17. Pilet, J., Geiger, A., Lagger, P., Lepetit, V., Fua, P.: An all-in-one solution to geometric and photometric calibration. In: IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR 2006, September 2006, pp. 69–78 (2006)
18. Openscenegraph, `http://www.openscenegraph.org`
19. Open computer vision library, `http://sourceforge.net/projects/opencvlibrary/`
20. Siftgpu, `http://cs.unc.edu/~ccwu/siftgpu/`

Personal Space Modeling for Human-Computer Interaction

Toshitaka Amaoka[1,3], Hamid Laga[2], Suguru Saito[3], and Masayuki Nakajima[3]

[1] Department of Information Science, Meisei University
amaoka@is.meisei-u.ac.jp
[2] Global Edge Institute, Tokyo Inst. of Technology
hamid@img.cs.titech.ac.jp
[3] Graduate School of Information Science and Engineering, Tokyo Inst. of Technology
{suguru,nakajima}@img.cs.titech.ac.jp

Abstract. In this paper we focus on the Personal Space (PS) as a non-verbal communication concept to build a new Human Computer Interaction. The analysis of people positions with respect to their PS gives an idea on the nature of their relationship. We propose to analyze and model the PS using Computer Vision (CV), and visualize it using Computer Graphics. For this purpose, we define the PS based on four parameters: distance between people, their face orientations, age, and gender. We automatically estimate the first two parameters from image sequences using CV technology, while the two other parameters are set manually. Finally, we calculate the two-dimensional relationship of multiple persons and visualize it as 3D contours in real-time. Our method can sense and visualize invisible and unconscious PS distributions and convey the spatial relationship of users by an intuitive visual representation. The results of this paper can be used for Human Computer Interaction in public spaces.

1 Introduction

Recent advances in Human Computer Interfaces (HCI) brought a new range of commercial products that aim at connecting the physical and the virtual worlds by allowing the user to communicate with the computer in a natural way.

The goal of this paper is to provide a human-computer interaction tool in which the computer detects and tracks the user's states and his relation with other users, then initiates actions based on this knowledge rather than simply responding to user commands. In this work we focus on the concept of Personal Space (PS) [6] which is a non-verbal and a non-contact communication channel. Everyone holds, preserves, updates this space, and reacts when it is violated by another person. In public spaces, for example, people implicitly interact with each other using the space around them. Psychologists also studied the existence

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 60–72, 2009.

of PS in virtual and cyber worlds such as Second Life [16,3]. They found that many users keep a PS around their avatar and behave in accordance with it in the virtual world. PS is also one factor that makes a virtual agent behave naturally and human like inside the virtual world [4]. It is also an important factor when simulating and analyzing the behavior of people in a crowd.

Based on this concept, we provide a new human-computer interaction framework. In a first step, we propose a mathematical model of the PS based on four parameters: the distance between people, their face orientations, age, and gender. The proposed model can be used for simulating the behavior of virtual agents and also for non-contact Human Computer Interaction. In both cases, the first two parameters can be estimated automatically while the two others are set manually since they are attributes of the personality. In the case of HCI application, we automatically estimate the distances between persons and their face orientations from image sequences using Computer Vision (CV) technology. Based on this model we calculate the two-dimensional relationship of multiple persons and visualize it as 3D contours in real-time.

The remaining parts of the paper are organized as follows; Section 1.1 reviews the related work. Section 1.2 outlines the main contributions of the paper. Section 2 details the mathematical model we propose for modeling the PS. Section 3 describes the system and algorithms we propose for estimating the parameters of the PS. Results are presented and discussed in section 4. We conclude in section 5.

1.1 Related Work

Existing Human Computer Interaction (HCI) technologies are mostly based on explicit user inputs acquired through devices such as keyboard, mouse, and joystick. There have been also many attempts to emulate these devices using cameras and speech so that the user will be able to use his hands for other purposes. Examples of such input devices include the camera mouse [1] and voice commands in home environments [13].

There is however no HCI technology that interprets the meaning of distances between people, neither the use of communication through space and distance. This concept is reflected in the notion of Personal Space (PS) which is a non-verbal communication and behavior.

The concept of Personal Space, since its introduction by Edward T.Hall [6,5] and the discussion by Robert Sommer [12], is studied and applied in many fields; In robotics, the PS is considered as a factor for selecting a communication method between a robot and a human [14]. It can be used for example to model of the intimacy of a robot to other users.

To simulate realistic behavior of virtual agents in virtual worlds, researchers try to apply the rule of real human behavior to the agents. There are many studies for simulating natural conversation and behavior of virtual agents. Rehm et al. [10] focused on the conversation and cultural difference to make natural behavior in virtual world. They treated the "first meeting scenario" to apply their model. Mancini et al. [7] proposed an Embodied Conversational Agents (ECAs) called Greta that are virtual embodied representations of humans that communicate multimodally with the user or other agents through voice, facial expression, gaze, gesture, and body movement. They studied what happens in real life, how people differ in their behavior depending on the personality and situation and build up their ECAs. Furthermore, many researchers study to make natural conversation and behavior to virtual agents as macro-level like group or crowds. McDonald et al. [8] proposed human behavior models (HBMs) that are able to control a single simulated entity (or a single group of simulated entities). However, HBMs developed by different groups are unable to interact with each other. This work mainly treats crowd control.

Quera et al. [9] models the space around an agent as a 2D function that reflects in each spatial location the degree of non-satisfaction of the agent if he moves to that location. The optimal configuration of a crowd of agents emerges from the local interactions. This is the closest model to our work. The non-satisfaction function they propose depends on inter-agent distances only. However, the PS concept is based on many complex rules. Its shape and size are affected by several factors such as gender, age, and social position in addition to distances and face orientations. In this paper, we propose a new model that takes into account all these factors.

1.2 Overview and Contributions

We propose a HCI framework based on the concept of PS to create interactive art. The idea is to interpret distances between people and make interaction with it. The proposed method estimates the shape of the personal space by measuring first the user's position and face orientation. Then we can simulate the way people communicate between each other through the interaction of their personal spaces. It generates and varies the contents according to the relationship between users.

The contributions of this paper are three-fold; first we propose a mathematical model of the PS. It is controlled by four parameters: the user's age, gender, position in the 2D space, and the face orientation. Second, we propose to estimate these parameters. User's position and face orientation are estimated automatically using a set of cameras mounted around the operation area. At the current stage the gender and age are set manually by the user. Finally, we visualize the PS in 3D using computer graphics.

The proposed method in this paper enables the detection of the user's mobile territory and his relationship with other. Our targeted application is interactive and collaborative digital art, but results of this work can be applied to modeling the behavior of virtual agents, as well as analyzing people behavior in a crowd.

2 Modeling the Personal Space

Edward T.Hall in his study of human behaviors in public spaces [6] found that every person holds unconsciously a mobile territory surrounding him like

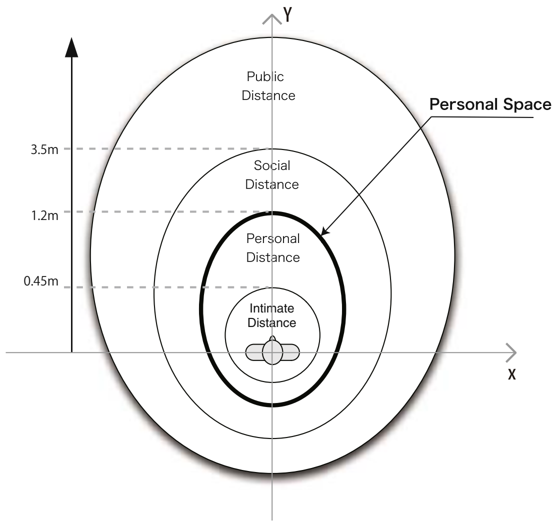


Fig. 1. Definition of the Personal Space. The space around a person is divided into four regions: the intimate distance, the personal distance, the social distance, and the public distance. The figure shows typical sizes of each zone.

Table 1. Standard personal space distances

	Distance	**Type of relationship**
Intimate distance	0 – 45cm	Very intimate relationship
Personal distance	45 – 120cm	Friends
Social distance	1.2 – 3.5m	Strangers
Public distance	> 3.5m	Public speaking

bubbles. The violation of this personal space by a tierce person results in an effective reaction depending on the relation of the two persons. This suggests that the concept of PS is a non-verbal communication between two or more persons. The personal space as defined by Edward T.Hall and shown in Fig. 1 is composed of four areas: the intimate space, the personal space, the social space and the public space.

The shape of the PS is affected by several parameters. In this paper we consider four of them: gender, age, distance, and face orientation. The relationship between gender and PS is well studied by sociologists [6]. They suggested also that the shape of the PS varies with the face orientation. For example, the PS is twice wider in the front area of a person than in the back and side areas.

2.1 The Model of the PS

Given a person P located at coordinates $p(x, y)$ we define a local coordinate system centered at p, with X axis along the face and Y axis along the sight direction as shown in Fig. 1. The personal space around the person P can then be defined as a function Φ_p which has its maximum at p and decreases as we get far from p. This can be represented by a two-dimensional Gaussian function Φ_p of covariance matrix Σ, and centered at p:

$$\Phi_p(q) = e^{-\frac{1}{2}(q-p)^t \Sigma^{-1}(q-p)}. \tag{1}$$

where Σ is a diagonal matrix:

$$\Sigma = \begin{pmatrix} \sigma_{xx}^2 & 0 \\ 0 & \sigma_{yy}^2 \end{pmatrix}. \tag{2}$$

The parameters σ_{xx} and σ_{yy} define the shape of the PS. Considering the fact that the PS is twice wider in front along the sight line than the side (left and right) areas, we define $\sigma_{yy} = 2\sigma_{xx}$.

This model assumes that the shape of the front and back areas of the PS are similar. However, previous studies pointed out that people are more strict regarding their frontal space. Shibuya [11] defines the PS in the front of people as twice larger as the back, left and right ares. We use this definition in our implementation. We build this model by blending two Gaussian functions as follows:

$$\Phi_p(q) = \delta(y_q)\Phi_p^1(q) + (1 - \delta(y_q))\Phi_p^2(q). \tag{3}$$

where $q = (x_q, y_q)^t$, $\delta(y) = 1$ if $y \geq 0$, and 0 otherwise. Φ_p^1 models the frontal area of the person and is defined as a $2D$ Gaussian function of covariance:

$$\Sigma_1 = \begin{pmatrix} \sigma_{xx}^2 & 0 \\ 0 & 4\sigma_{xx}^2 \end{pmatrix}. \tag{4}$$

Φ_p^2 models the back area of the person and is defined as a $2D$ Gaussian function of covariance

$$\Sigma_2 = \begin{pmatrix} \sigma_{xx}^2 & 0 \\ 0 & \sigma_{xx}^2 \end{pmatrix}. \tag{5}$$

Notice that the standard deviation of Φ_p^1 along the Y axis is twice the standard deviation of Φ_p^2 along the same axis. The function δ blends the two functions and therefore it allows to take into account the face orientation. This concept is illustrated in Fig 2. In our implementation, we model the space where people (or agents) are interacting as a 2D plane parallel to the floor plane. We define a unique world coordinate system and encode a the floor plane as a 2D matrix. Every person P_i holds such a matrix (herein after referred as M_i). Each element of the matrix encodes the importance of the corresponding location to the person. The matrix is dynamic and is updated every time step δt. Fig. 1(b) shows the variation of this matrix according to location and face orientation.

In our implementation we define σ_{xx} as the threshold to which a specific zone of the space is violated. For example, to model the intimate space of a standard

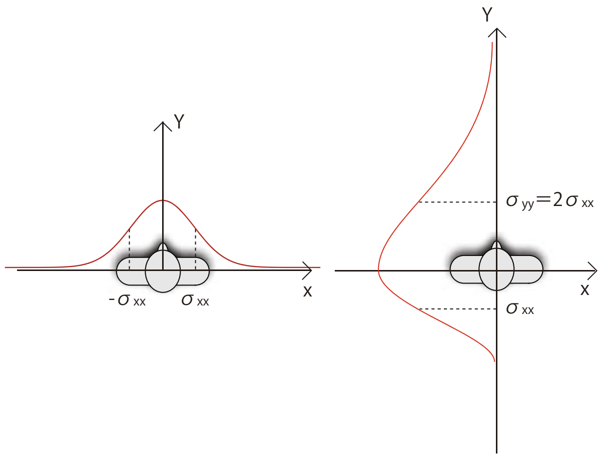


Fig. 2. The Personal Space model based on the face orientation. The PS in the front area of a person is wider that the back and side areas.

person we set $\sigma_{xx} = \sigma_0 = 0.45/2 = 0.255m$ as shown in Fig. 1. The figure gives also the standard values of σ_{xx} for different zones of the PS.

2.2 Parameterization of the PS

The personal space does not depend only on the position and face orientation but also on factors related to the person such as age, gender, social position, and character. These personal factors can be included in a function f that affects the value of the standard deviation σ_{xx}. In our implementation we consider only age and gender, hence:

$$\sigma_{xx} = f(\sigma_0, age, gender). \tag{6}$$

In the simplest case, f can be a linear function that scales σ_0 with a factor α reflecting how much a person is kin to protect his intimate space. This model is however not realistic. In our implementation we encode the age and gender dependency as a lookup table where each entry corresponds to the value of σ_{xx} given the age and gender. Table 2 shows an example of such table as defined in [2]. The distances are given in centimeters, and are for Anglo ethnic group. The table also shows that the personal space varies depending on the situation such as being in indoor or outdoor environment. In summary the Personal Space model is parameterized by two types of factors:

- The inter-personal factors, which include distances and face orientations, are embedded inside the parameter σ_{xx}. These two parameters are estimated automatically as will be explained in Section 3.
- The personal parameters such as age, gender, and social position embedded in the function f. These parameters are input manually by the user and are encoded as a lookup table.

Table 2. Variation of the frontal area of the personal space with respect to age and gender. The values vary with ethnicity. We considered only Anglo ethnic group in indoor and outdoor locations [2]. The distances are in *cm* and are equivalent to $2\sigma_{xx}$.

Sex combination	Indoor			Outdoor		
	Adult	Teenage	Child	Adult	Teenage	Child
M-M	83	83	63	83	75	62
M-F	71	63	58	79	68	58
F-F	75	62	59	75	73	67

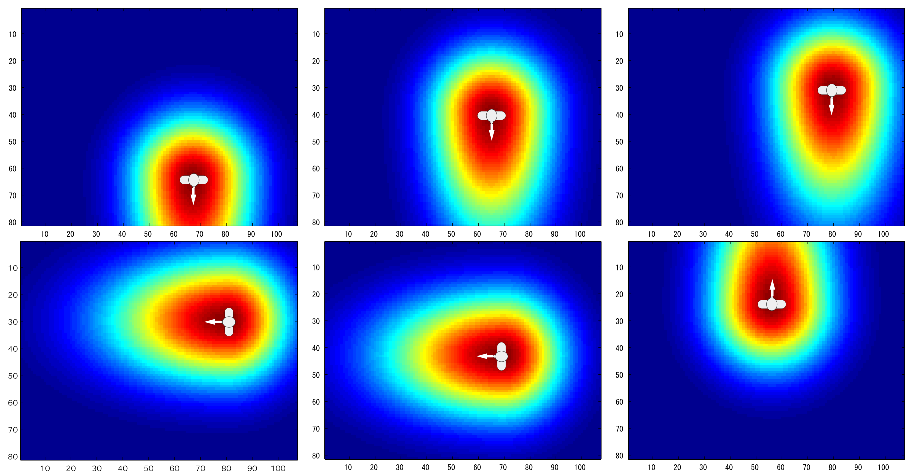


Fig. 3. Illustration of the representation of the PS function as a 2D matrix where each cell encodes the importance of the corresponding location to the agent / person. The arrow indicates the direction of the face.

In the following we describe the computer vision platform we developed for estimating the inter-personal factors used for simulating the Personal Space.

2.3 Implementation Details

In our model, the personal space is computed with respect to the floor (XY) plane. We represent the space as a 2D matrix where each cell (i, j) represents a floor location. Each person keeps and updates every time step Δt his own matrix, herein referred as M_a. It encodes in each cell the importance of the corresponding location which is computed using Equations 3 and 6.

Figure 3 shows how the matrix is updated with respect to the agent (or person) location and face orientation.

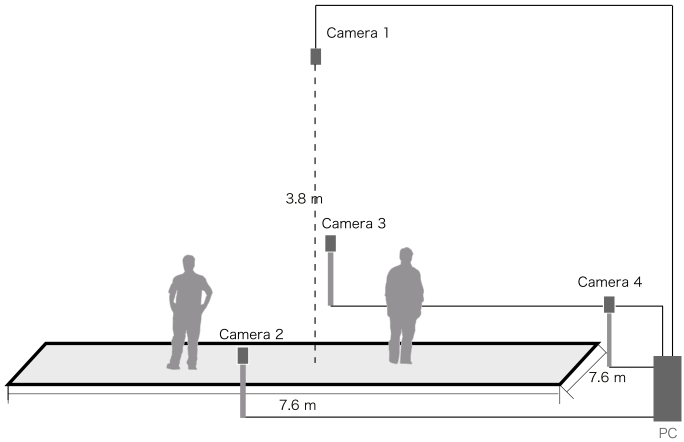


Fig. 4. Overview of the system setup. The top camera (camera 1) is used for people detection. The side cameras (camera 2-4) are used for detecting faces and estimating their orientation.

3 System

Using the model of Eq. 3 for interaction requires the estimation of the user's position and face orientation. We build a computer vision platform to track the user's behavior in front of the screen. Using a 3D sensor for efficient localization and tracking is possible and will provide more accurate data. However, 3D sensors such stereo cameras and laser range scanners are expensive. The system, as shown in Figure 4, is composed of four IEEE1394 cameras: one overhead camera for people detection and tracking, and three frontal cameras for detecting faces and estimating their orientation.

3.1 Multiple User Tracking

We detect people by detecting and tracking moving blobs in the scene. For efficient detection and tracking, we setup the system in a studio under controlled lighting conditions. The system then operates as follows:

- First, using the overhead camera, we take shots of the empty scene and use them to build a background model by learning a Mixture of Gaussian classifier. Given a new image I, the classifier classifies every pixel into a background or foreground.
- The persons are detected as connected components of the foreground pixels.
- Second, during the operation, we track people over frames using the Mean Shift algorithm.

Estimating distances between people requires recovering the 3D location of the detected persons. To do so, we setup a world coordinate system in such a way

Camera 1 Camera 2 Camera 3 Camera 4

Fig. 5. Example of person and face detection and tracking using four cameras

that the XY plane corresponds to the floor plane, and the Z plane oriented to the upper direction. The cameras are calibrated and therefore, the exact locations of the persons. We use the face location as the location of the persons. We explain how faces are detected in the next subsection.

3.2 Face Detection and Orientation Estimation

To estimate the face orientation in the 3D space we

- Detect faces from each of the three frontal cameras using the Viola and Jones face detector [15] freely available with the Intel's OpenCV library.
- Assign the detected facial images to each of the persons detected with the overhead camera.

Since the entire camera system is calibrated, i.e., the transformations matrices between the frontal cameras and between each frontal camera and the top one are computed in advance, the faces are assigned to the persons by assigning each face to its closest blob in the overhead camera.

In our current implementation, we are considering only four orientations of the faces corresponding to the angles $0, 90, 180, 270$ degrees. When the a person is facing camera 2 (see Fig. 4), only camera 2 detects the frontal face. Therefore, we assign the angle 0. We apply the same procedure to the other cameras. Camera 4 corresponds to angle 90 and camera 3 corresponds to angle 180. We assume the orientation 270 when no face has been detected by the three frontal cameras.

Figure 5 shows an example where people positions and faces are detected using our system.

4 Results

To visualize the Personal Space of a person we consider the three levels as shown on Fig. 1. We define a PS function for each zone using Equation 3. Position and face orientation are estimated automatically.

The PS function as defined in Equation 3 can be also interpreted as the degree of the user's response to the violation of a zone in his PS. Depending on his relation with the other person, the user activates one of the three functions.

Fig. 6. Visualization of the personal area of the personal space. The user is facing Camera 3 (X axis).

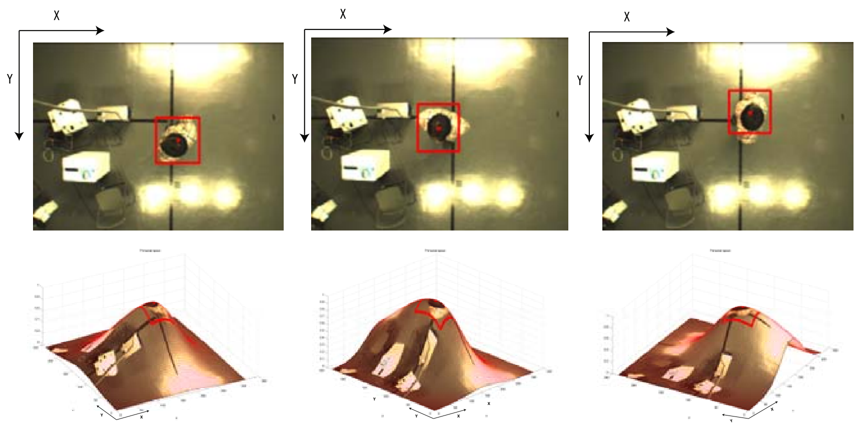


Fig. 7. Visualization of the PS when the face is rotating from camera 2 to camera 4

To visualize this, we map the image space (taken by the top camera) onto the XY (the floor) plane of the world coordinate system. Since we used only the frontal face detector, we detect three orientations of the face $0, 90$ and 180 degrees. Figure 6 shows an example where the user is facing Camera 3. Notice that camera 3 has detected the face and therefore the area of the PS in front of the user is larger than back and side areas.

Figure 7 shows an example where the user is standing and rotating only his face from one camera to another. Notice how the personal space evolves with the face orientation. This particularly proves the efficiency of our model.

Finally, Figure 8 shows two persons in the same scene and their associated personal space matrices. Each pixel in the matrix encodes how important is the corresponding location to the person. This figure also shows how the PS shape varies according to the face orientation. In our current implementation we considered only the four principal orientations $(0, 90, 180, 270)$. Other orientation can be considered by improving the face detection and orientation estimation part.

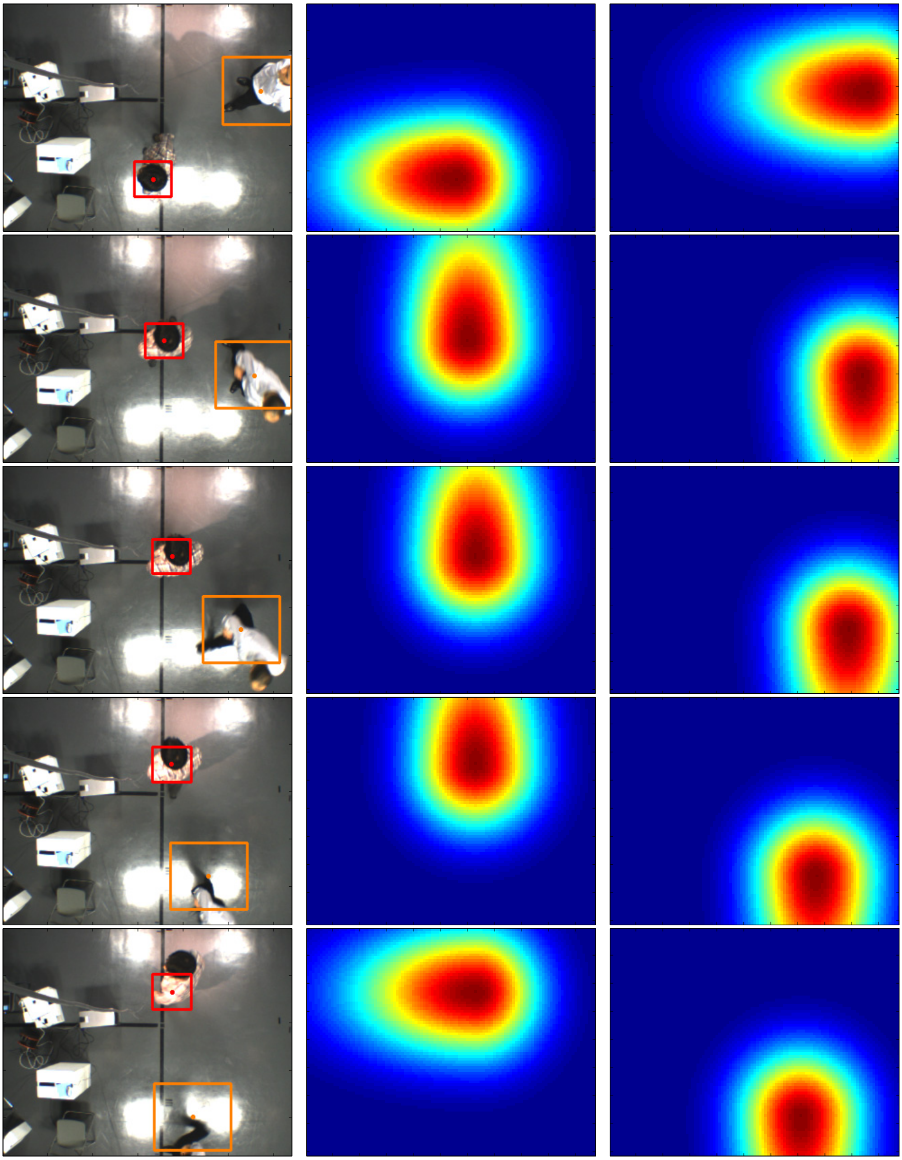


Fig. 8. Variation of the PS with respect to distances and face orientations when two users are interacting. The middle column corresponds to the PS of the person in read.

5 Conclusion

In this paper we have proposed a mathematical model of the personal space. We particularly implemented a method for automatically estimating two parameters of the PS: the people position and their face orientation. As extension we plan in future to automate the estimation of the other parameters such as age and gender. Possible extensions include also the improvement of face detection algorithms for detecting faces at different orientations. We plan also in the future to use this personal space model for HCI and also modeling virtual agents behavior.

Acknowledgement. Hamid Laga is supported by the Japanese Ministry of Education, Culture, Sports, Science and Technology program Promotion of Environmental Improvement for Independence of Young Researchers under the Special Coordination Funds for Promoting Science and Technology.

References

1. Camera mouse, http://www.cameramouse.org/ (accessed on November 15, 2008)
2. Baxter, J.C.: Interpersonal spacing in natural settings. Sociometry 33(4), 444–456 (1970)
3. Friedman, D., Steed, A., Slater, M.: Spatial social behavior in second life. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 252–263. Springer, Heidelberg (2007)
4. Gillies, M., Slater, M.: Non-verbal communication for correlational characters. In: Proceedings of the 8th Annual International Workshop on Presence, London (September 2005)
5. Hall, E.T.: Proxemics. Current Anthropology 9(2-3), 83–108 (1968)
6. Hall, E.T.: The Hidden Dimension: Man's use of Space in Public and Private. Anchor Books, Re-issue (1990)
7. Mancini, M., Hartmann, B., Pelachaud, C.: Non-verbal behaviors expressivity and their representation (2004)
8. McDonald, D., Lazarus, R., Leung, A., Hussain, T., Bharathy, G., Eidelson, R.J., Pelechano, N., Sandhaus, E., Silverman, B.G.: Interoperable human behavior models for simulations (2006)
9. Quera, V., Beltran, F., Solanas, A., Salafranca, L., Herrando, S.: A dynamic model for inter-agent distances. In: Meyer, J.A., Berthoz, A., Floreano, D., Roitblat, H.L., Wilson, S.W. (eds.) From animals to animats, The International Society for Adaptive Behavior, vol. 6, pp. 304–313 (2000)
10. Rehm, M., Nakano, Y., Andres, E., Nishida, T.: Culture-specific first meeting encounters between virtual agents, pp. 223–236 (2008)
11. Shozo, S.: Comfortable distance between people: Personal Space. Japan Broadcast Publishing Co., Ltd. (1990)
12. Sommer, R.: Personal Space: The Behavioral Basis of Design. Prentice Hall Trade, Englewood Cliffs (1969)
13. Soronen, H., Turunen, M., Hakulinen, J.: Voice commands in home environment - a consumer survey - evaluation (September 2008)

14. Tasaki, T., Matsumoto, S., Ohba, H., Yamamoto, S., Toda, M., Komatani, K., Ogata, T., Okuno, H.G.: Dynamic communication of humanoid robot with multiple people based on interaction distance. Transactions of the Japanese Society of Artificial Intelligence 20, 209–219 (2005)
15. Viola, P., Jones, M.: Robust real-time face detection. International Journal of Computer Vision 55(2), 137–154 (2004)
16. Yee, N., Bailenson, J.N., Urbanek, M., Chang, F., Merget, D.: The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. Journal of CyberPsychology and Behavior 10, 115–121 (2007)

Technology-Enhanced Role-Play for Intercultural Learning Contexts

Mei Yii Lim[1], Michael Kriegel[1], Ruth Aylett[1], Sibylle Enz[2], Natalie Vannini[3], Lynne Hall[4], Paola Rizzo[5], and Karin Leichtenstern[6]

[1] School of Mathematical and Computer Sciences, Heriot Watt University, Edinburgh, EH14 4AS, Scotland
{myl,michael,ruth}@macs.hw.ac.uk
[2] Otto-Friedrich-Universität Bamberg, Kapuzinerstrasse 16, D-96045 Bamberg, Germany
sibylle.enz@uni-bamberg.de
[3] Universität Würzburg, Lehrstuhl für Psychologie IV Röntgenring 10, D-97070 Würzburg, Germany
natalie.vannini@psychologie.uni-wuerzburg.de
[4] School of Computing and Technology, University of Sunderland
lynne.hall@sunderland.ac.uk
[5] Interagens s.r.l., Via G. Peroni 444, 00131 Rome, Italy
p.rizzo@interagens.com
[6] Universität Augsburg, Lehrstuhl für Multimedia-Konzepte und Anwendungen Eichleitnerstr. 30, 86159 Augsburg, Germany
karin.leichtenstern@informatik.uni-augsburg.de

Abstract. Role-play can be a powerful educational tool, especially when dealing with social or ethical issues. However while other types of education activity have been routinely technology-enhanced for some time, the specific problems of supporting educational role-play with technology have only begun to be tackled recently. Within the eCIRCUS project we have designed a framework for technology-enhanced role-play with the aim of educating adolescents about intercultural empathy. This work was influenced by related fields such as intelligent virtual agents, interactive narrative and pervasive games. In this paper we will describe the different components of our role-play technology by means of a prototype implementation of this technology, the ORIENT showcase. Furthermore we will present some preliminary results of our first evaluation trials of ORIENT.

1 Introduction

Drama and play have been used for education for a very long time [1] and have resulted in game-based educational approaches. These provide a means of overcoming real-world social restrictions, placing the player in a role that may or may not be socially acceptable in real life, such as a medical doctor or a thief. Games allow the player to escape into fantasy worlds, encourage exploration of exciting things, people, and places that are otherwise inaccessible in the real

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 73–84, 2009.

world, inducing a 'suspension of disbelief' in the player. Learning often takes place while the game is played, with immediate feedback. The subject to be learned is directly related to the game environment where constant cycles of hypothesis formulation, testing and revision are evoked as the player experiences continuous cycles of cognitive disequilibrium and resolution.

This paper explores an approach to an educational role-play (RP) game developed in the ORIENT showcase of the eCIRCUS[1] project, employing innovative technologies to foster social and emotional learning in the adolescent age group. With globalisation, dealing with cultural difference and diversity has become a widespread task and is both challenging and enriching. In looking for ways to help the process of acculturation of adolescents from emigrant backgrounds, there were a number of reasons for not focusing on them directly. Firstly, they form a heterogeneous group with a multitude of cultures and languages. It would be infeasible to try to capture all these in a computer-based system. More than this, acculturation is a two-way process in which both the incoming group and the host group have to negotiate a common understanding. It was therefore decided to focus on the host group, and to foster intercultural sensitivity through the development of intercultural empathy.

By increasing the social and intercultural competence of the host adolescents, ORIENT aims at diminishing discrimination and hence lowering the mental stress of peers from a migration background. ORIENT offers a virtual role-play environment inhabited by autonomous artificial agents that interact with and react to a group of learners. Hence, learners may collaboratively improve their perception of and alter their emotional reactions and attitudes to members of other cultures in a secure social setting, while interacting with the virtual environment through a set of engaging and immersive interaction devices.

2 Related Work

Pervasive gaming takes virtual narrative elements out into the real world, focusing on introducing game elements into the everyday life of players. They exploit interaction devices such as handhelds to display virtual world elements [2] and employ technology support through which human game-masters can exercise higher amounts of control over the game experience [3]. The Enhanced Reality Live Role-Playing of the IPerG project, in the area of pervasive games, has successfully carried out a number of pervasive games in real spaces (e.g. [4]). These focused on the idea of linking the real world into the story world [5], through for example, using unwitting inhabitants of the real world as props for pervasive game players. Some other groups have also produced educational pervasive games. Virus [6] is a game in which learners take on the role of a virus and transmit it via specially-designed mobile devices called Thinking Tags by getting within proximity of other users. This demonstrated a complex disease-propagation algorithm in a real world setting. In Paranoia syndrome [7], learners can take on the roles and skills of a technician, doctor or scientist. The Virtual

[1] http://www.e-circus.org/

Savannah [2] took child learners out of the classroom setting and through the use of handheld devices made it possible for them to view their school playing field as a Savannah on which they role-played lions. A more recent and more problem-oriented role-play, the Environmental Detective [8] used a whole university campus as its story-world, while artistically oriented pervasive games such as Uncle Roy All Around You and I Can See You Now [2] have used whole cities as the game environment.

The use of large-scale real-world spaces for role-play suits some applications, but others require a dedicated space, and can be thought of as stage-based role-play. This is true of many of the existing educational role-plays. A stage-based environment can be thought of as a sensor-rich pervasive computing environment including large display systems in which virtual actors and graphical worlds can play a more prominent role than is feasible when only hand-held devices are used. An early example of the stage-based approach is the Mission Rehearsal Exercise [9] in which a single human participant interacts with virtual characters in a stressful and dramatic situation (peacekeeping) using structured speech. This work has been extended into a more augmented reality environment using 'flats' - large display screens within a real world space, but has limited interaction modalities. ORIENT also takes this approach for a role-play that aims to educate students in inter-cultural empathy employing a set of innovative interaction devices.

3 The Game: ORIENT

3.1 The Story

ORIENT is being developed for the 13-14 age group of boys and girls in the UK and Germany and is designed to be played by a group of 3 teenage users. Each one of them takes the role of a member of a spaceship crew. Their mission takes them to a small planet called ORIENT, which is inhabited by an alien race - the nature loving Sprytes. Portraying a fictional instead of an existing culture makes our application more flexible and suitable for users from diverse backgrounds. Furthermore, it allows us to exaggerate cultural differences for dramatic and educational purposes.

The Sprytes are not aware of the danger that their planet is in: a meteorite is on destruction course and unless someone stops it, it would mean the end of life on ORIENT. It is the users' task to prevent a catastrophe. To do that the users first have to befriend the Sprytes and ultimately cooperate with them to save their planet. Through interaction with the Sprytes, ORIENT promotes cultural-awareness in the users, who have to put themselves into the shoes of guests to a strange and unknown culture. At the same time ORIENT acts as a team building exercise where users play as a single entity rather than as individuals. All users have the same goal in the game although their roles and capabilities differ.

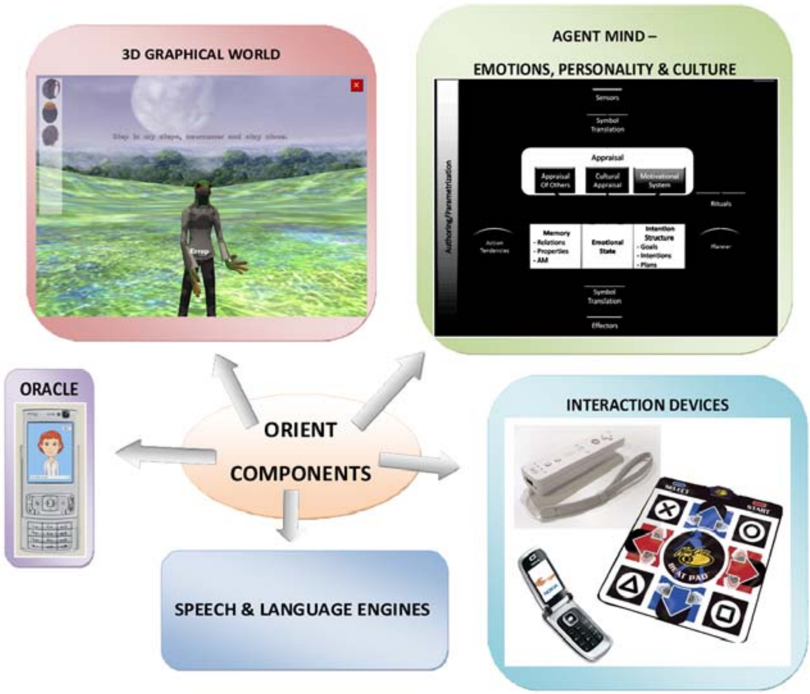


Fig. 1. ORIENT system components

3.2 The ORIENT Prototype

A prototype of ORIENT has been implemented consisting of the main components shown in Figure 1. Each component will be described separately in the next section. In this prototype a group of 3 users explores 4 different locations of the Sprytes' world . All users share a single first person perspective of the same 3D virtual world. Each user is equipped with a different interaction device (Section 4.3), all 3 of which are necessary to interact with the virtual Spryte characters. A Spryte character is implemented as an autonomous agent based on the FAtiMA [10] agent mind architecture (Section 4.1). Dialogues are treated by the agent mind as symbolic speech acts. When a Spryte speaks, a language engine transforms the speech act into natural language subtitles while a speech engine simultaneously generates the respective audio in an artifcial incomprehensible gibberish language. While the application is running, both the ORACLE (Section 4.2) and the Story Facilitator modules are constantly monitoring all events in the game world. The ORACLE uses this information to provide context-sensitive help and advice to the users. The Story Facilitator on the other hand, monitors the events in order to ensure an interesting story develops. This is achieved by directly influencing the game world, for example by introducing a new character in certain situations.

3.3 Culture in ORIENT

According to Kluckhohn [11], culture is that part of behaviour which is learned by people as the result of belonging to some particular group. “It is the main factor which permits us to live together in a society, giving us ready made solutions to our problems, helping us to predict the behavior of others, and permitting others to know what to expect of us” (p. 25).

Classifications of cultures and a taxonomy on dimensions of cultural variability have been provided by authors including Hall [12] and Hofstede [13]. Hofstede defines the following dimensions of cultural variability: individualism-collectivism; uncertainty avoidance; power distance; and masculinity-femininity. These have been used to define the culture of the Sprytes. This is a tribal culture with a hierarchy (high power distance), depending highly on respect and age. This is reinforced by the fact that Sprytes are militarily active and believe in using force and power to influence others and to protect their habitat. They are a collectivistic culture (individualism-collectivism), which makes them compassionate

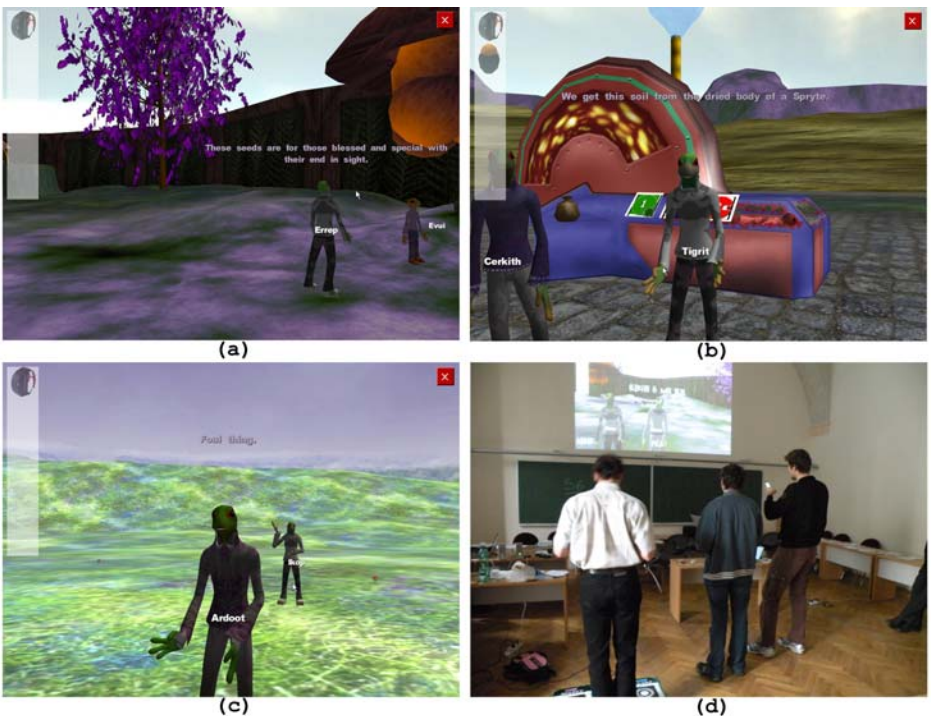


Fig. 2. (a) Educating a child Spryte for picking seedpod from the tree which is against the Spryte’s culture; (b) A Spryte explaining their life cycle; (c) Angry gesture to the user for stepping on a little tree; (d) users interacting with ORIENT

with each other, and live in groups where the majority holds power. They are highly traditional in their ways and view uncertainty as a threat (uncertainty avoidance) but exceptions do exist in younger Sprytes. Gender (masculinity-femininity) is absent from the Sprytes and the graphical representation of the Sprytes is intended to be ambiguous from this point of view as can be observed in Figure 2(a-c).

4 System Components

4.1 Virtual Actors

The use of virtual actors is one of the most important ways of shaping the narrative experience in RP games. In order to produce competent and responsive virtual actors, ORIENT draws upon previous work in AI in the fields of affective-cognitive models, intelligent synthetic characters and embodied conversational agents (ECAs), and interactive narrative. Affective models are seen here as a key component of virtual actors given the need to affectively engage participants and spectators in the dramatic environment. These were absent from early work in virtual actors, as in IMPROV [14] and the Virtual Theater project [15] in which virtual actors operated purely within graphics environments and were at most semi-autonomous.

The ORIENT software is being built upon FAtiMA [10], an agent architecture with an emotional continuous planner. We have extended FAtiMA with concepts from the PSI [16, 17] model, which bases all actions on the fulfilment of basic physiological needs. Here, we take the body-mind [18] view of emotion where emotions do not rely purely on reasons but there exists an interaction between physiological processes and the cognitive processes in a human action regulation system.

Employing FAtiMA, the agent's decision making processes are influenced by the OCC [19] cognitive appraisal. OCC is a widely used taxonomy for categorising and explaining emotional occurrences. The advantage of using the OCC model for ORIENT characters is that empathy can be modelled easily because it directly relates to the appraisal of events as they impact on others. It is - as far as we know - the only model that provides a formal description of non-parallel affective empathic outcomes. Additionally, the OCC model includes emotions that concern behavioural standards and social relationships based on like/dislike, praiseworthiness and desirability for others, allowing appraisal processes that take into consideration cultural and social aspects, important for the believability of ORIENT characters. Characters are entirely 'in-role' in that they select actions according to their immediate goals and environment, portraying affective engagement with their own situation and each other. This lays the basis for affective engagement of the learners with the characters.

The PSI model extends the empathic modelling to more emotional outcomes than those described by the OCC model. Emotions within the PSI model are conceptualised as specific modulations of cognitive and motivational processes. The motivational system serves as a quick adaptation mechanism of the agent to a specific situation and may lead to a change of belief about another agent [20], important for conflict resolution among ORIENT characters. Utilising PSI, processes in ORIENT characters become self-regulatory and parallel, driven by needs that emerge from activities in the environment or grow over time. Each character will continuously perceive the environment and create intentions that may satisfy its needs at a particular instance of time. One of these intentions is then selected for execution based on the degree it satisfies the character's needs and its probability of success. Through trial and error by executing different goals, ORIENT characters learn. The character's experiences are stored in an autobiographic memory [21] for future reference, thus, permitting adaptive and flexible behaviour in the dynamic RP environment so that the RP is open-ended rather than pre-scripted. For more information on the ORIENT architecture, please refer to [22].

4.2 The ORACLE

The ORACLE (Onboard Resource Agent for Cultural and Liaison Engagements) is an embodied computer character that aims at enhancing user's learning in the game. The ORACLE as shown in Figure 1 runs on a Nokia N95 phone. It plays the role of a human facilitator in fostering users' motivation and keeping them engaged, stimulating group collaboration, keeping the users' focus on the task and providing help during the mission. Its advice focuses mainly on facilitating and stimulating the intercultural learning processes but also includes more pragmatic help with the technology.

The ORACLE stimulates users' reflection on events and outcomes of ORIENT by asking appropriate questions and commenting on users' actions. Help is provided as hints to the team when it notices that the team is not making any progress. Users can also ask questions by selecting buttons on the graphical user interface. The ORACLE mind mainly consists of a production system containing two types of rules: "reactive" ones, that fire when the user asks for help, and "proactive" ones, that fire according to the occurrence of specific events in ORIENT.

4.3 Interaction Modalities

Approaches such as tangible user interfaces, mobile user interfaces and augmented reality all attempt to combine the physical world of users and the virtual world of applications to make interactions easier, quicker, more fun and more engaging for both single and multi-user interactions. Including the real world in the interaction process offers new possibilities in performing computer supported tasks. Research identifies that increasing the integration of the user's physical environment will render interaction more intuitive and engaging (eg. [23, 24]). Innovative interaction devices that provide more intuitive interfaces to express affect has been shown to enhance empathic engagement [25]. Additionally, empirical evidence is found that a more positive impact on collaboration can be achieved in learners when using a computer with multiple input devices and cursors than when using one without these interaction facilities (eg. [26, 27, 28]).

It was observed that by assigning each user a specific role tied to an interaction device with a dedicated function, more organised interaction within a group is produced, balancing the level of interactivity and avoiding dominant users [28].

Taking these studies as guidelines, ORIENT's user interface was designed to be physical and tangible reducing the discrepancy between action and perception. Interaction is supported through large and micro screens, physical interfaces and multi-modal interaction devices. Full body interaction and movement in the physical space, particularly important in social behaviour and culturally specific interaction are supported (see Figure 2(d)). Each user is assigned a role which relates to a specific interaction device - a mobile phone, a Dance Mat or a WiiMote - that has unique functions, necessary to achieve the overall goal of the game.

The Nokia NFC 6131 mobile phone supports RFID-based input and speech recognition. The user who is assigned to use the phone selects objects by touching RFID tags-embedded objects that also have existence in the virtual world. To grab a virtual character's attention, he calls out a "magic word" (the character's name).

A second user uses a Dance Mat to navigate the virtual world. This interface has two advantages: first it bears a resemblance to real-world navigation in the sense that the user has to take actual steps and second since it only occupies the user's feet, it allows this user to operate the ORACLE at the same time.

The WiiMote is used for three-dimensional gesture recognition based on motion data derived from accelerometer sensors. We use the Wiigle library for this purpose [29]. The third user performs different gestures using the WiiMote as a communication channel with the Sprytes. The use of gestures for communication eliminates the need for natural language processing.

5 Interaction Scenario

During the mission, the users will witness the Sprytes eating habits - eating only seedpods that have dropped onto the ground (Figure 2(a)), life cycles - recycling the dead (Figure 2(b)), educational styles, family formation and value system - trees are sacred (Figure 2(c)). An example scenario that is related to the Sprytes' eating habit is described below:

The interaction starts with the users greeting the Sprytes (performing the greeting gesture using the WiiMote). Then, the users witness a Spryte picking a seedpod from the ground and eating it. On the other hand, there will be a child Spryte who picks a seedpod from the tree. Once the child picked a seedpod from the tree, an adult Spryte will start the education process (because picking or eating seedpods from trees is forbidden in the Spryte culture) from which the users can learn about the Sprytes' culture. If the users approach a tree (stepping forward on the Dance Mat in the direction of a tree), they will be warned by one of the Sprytes about their inappropriate behaviour. If the warning is ignored and the users pick a seedpod from the tree (perform the pick gesture using the WiiMote and scan a RFID-embedded seedpod), the Sprytes will be angry with the *users. A friendly Spryte might put in a good word for the users with his angry mate in this case. At this point, the Story Facilitator might interrupt by creating an external event, for example, a meteor hitting one of the Sprytes. The users can then act with the help of the ORACLE to cure the dying Spryte and by doing so, achieve redemption. If the users pick a seedpod from the ground, they will be invited for a meal together with the Sprytes. The users can choose to accept or reject the invitation (performing accept or reject gesture using the WiiMote) and their response affects future relationship with the Sprytes.*

In the above scenario, the users have possibilities of making "right" or "wrong" choices thus highlighting the conflict management aspects of multi-cultural integration. Through direct feedback from the interaction, the users can analyse the appropriateness or suitability of their action. This coupled with post-role-play reflection or debriefing will help the users to transfer their experiences to real-world settings, hence, learn to adapt and accept differences among cultural groups.

6 Evaluation and Discussion

The ORIENT prototype has been evaluated in a lab-based, small-scale study in the UK and Germany with a total of 12 adolescents respectively. Each evaluation session took approximately 2 hours (including device training, actual ORIENT interaction and filling out the questionnaires). The key aim is to test the suitability of ORIENT as a tool for:- (a) fostering cooperation/collaboration; and (b) fostering reflection on intercultural problems. The evaluation framework is placed within the RP scenario itself, making the evaluation part of the interaction rather than a separate experience. The users operated as members of "Space Command" throughout the evaluation. To support this in-role approach, questionnaires and training videos were designed in a "Space Command" style and our evaluation team members were acting according to predefined roles. The two hypotheses were evaluated using post-interaction questionnaires (Cultural Intelligence Scale (CIS, [30]); Intergroup Anxiety Scale [31]; General Evaluation Scale [32]) and by interaction observation (Team cooperation, Solution-orientation, Empathy, Egocentrism, Small-Talk, Integration, Interaction, and Group mood).

The initial results indicate that the prototype and the use of the different interaction devices have rudimentarily the potential to foster cooperation among the user group. The groups reported quite positive feedback regarding their interaction. One group expressed that they had the feeling they had played very well together as a team. The participants were capable to point out similarities and differences between their own and the culture of the Sprytes and expressed rather positive feeling towards the Sprytes.

The technical evaluation focused on the experience of interacting with ORIENT (ORIENT Evaluation Questionnaire), the usability of the ORACLE (ORACLE Evaluation Form), and on the usability of the interaction devices (Device Questionnaire). The participants found it interesting to handle the different devices, and that all devices were needed to accomplish the interaction with the Sprytes despite the fact that it took them quite a while to be able to control the devices.

Some issues were discovered during the evaluation including the Sprytes' personality, interaction devices and the story content. It was found that believability as well as the emotional impact of the drama on the users could be fostered by giving the individual Sprytes more distinctive personalities as they currently appear as cultural stereotypes. In order to enable the users to explore, understand, and flexibly react to the cultural encounter with and the specific problems of the Spryte culture, the interaction with the virtual world should be even more intuitive and seamless: sometimes the users found it hard to perform certain WiiMote gestures or pronounce characters' names correctly. Furthermore, the software needs to be enhanced content-wise in order to make the encounter with the Spryte culture a pedagogically meaningful experience, for example, adding content that may lead to some kind of conflict with the Sprytes, either a vicariously experienced conflict located within the Spryte culture, or a "real" conflict between the culture of the Sprytes and the user group.

7 Conclusion

The ORIENT software provides a role-play and story-framework for virtual social actors to interact with users in a number of different ways so as to create inter-cultural empathy. It employs tangible interaction modalities to increase users' motivation to learn about the Sprytes' culture and their engagement in the interaction, at the same time to enhance collaboration among themselves. It exhibits the potential of technology-enhanced role-play to support social and emotional learning in complex social situations without the risks that the learner faces in a real social situation due to the secure settings for experimentation with new behavioural strategies.

Acknowledgements

This work was partially supported by European Commission (EC) and was funded by the eCIRCUS project IST-4-027656-STP with university partners Heriot-Watt, Hertfordshire, Sunderland, Warwick, Bamberg, Augsburg, Würzburg plus INESC-ID and Interagens. The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

References

[1] Crawford, C.: The Art of Computer Game Design, Chapter 2: Why Do People Play Games? Washington State University (1997), http://www.vancouver.wsu.edu/fac/peabody/game-book/Chapter2.html

[2] Benford, S., Crabtree, A., Flintham, M., Drozd, A., Anastasi, R., Paxton, M., Tandavanitj, N., Adams, M., Row-Farr, J.: Can you see me now? ACM Trans. Comput.-Hum. Interact. 13(1), 100–133 (2006)

[3] Söderberg, J., Waern, A., Åkesson, K.P., Björk, S., Falk, J.: Enhanced reality live role playing. In: Workshop on Gaming Applications in Pervasive Computing Environments, Second International Conference on Pervasive Computing, Vienna, Austria (2004)
[4] Lindt, I., Ohlenburg, J., Pankoke-Babatz, U., Ghellal, S.: A report on the cross-media game epidemic menace. In: Computers in Entertainment (2007)
[5] Stenros, J., Montola, M., Waern, A., Jonsson, S.: Play it for real: Sustained seamless life/game merger in momentum. In: Proceedings of DiGRA 2007 Situated Play conference, Tokyo, pp. 121–129 (2007)
[6] Collella, V., Bororvoy, R., Resnick, M.: Participatory simulations: Using computational objects to learn about dynamic systems. In: SIGCHI conference on Human factors in computing systems (CHI 1998), Los Angeles, USA (1998)
[7] Heumer, G., Carlson, D., Kaligiri, S.H., Maheshwari, S., Hasan, W.U., Jung, B., Schrader, A.: Paranoia syndrome - a pervasive multiplayer game using pdas, rfid, and tangible objects. In: Third International Workshop on Pervasive Gaming Applications on Pervasive Computing 2006, Dublin, Ireland (2006)
[8] Klopfer, E., Squire, K.: Environmental detectives: the development of an augmented reality platform for environmental simulations. In: Educational Technology Research and Development, pp. 203–228 (2008)
[9] Hill, R., Gratch, J., Marsella, S., Rickel, J., Swartout, W.: Virtual humans in the mission rehearsal exercise system. In: Künstliche Intelligenz (2003)
[10] Dias, J., Paiva, A.: Feeling and reasoning: A computational model for emotional characters. In: Bento, C., Cardoso, A., Dias, G. (eds.) EPIA 2005. LNCS (LNAI), vol. 3808, pp. 127–140. Springer, Heidelberg (2005)
[11] Kluckhohn, C.: Culture and Behavior. The Free Press of Glencoe, New York (1988)
[12] Hall, E.T.: Understanding Cultural Differences. Intercultural Press, Yarmouth (1988)
[13] Hofstede, G.: Cultures and Organisations. McGraw-Hill, London (1991)
[14] Perlin, K., Goldberg, A.: Improv: A system for scripting interactive actors in virtual worlds computer graphics. In: International Conference on Computer Graphics and Interactive Techniques, pp. 205–216. ACM, New York (2006)
[15] Rousseau, D., Hayes-Roth, B.: A social-psychological model for synthetic actors. In: Proceedings of the 2nd International Conference on Autonomous Agents (1998)
[16] Dörner, D.: Bauplan für eine Seele. Rowohlt Taschenbuch, Reinbek (2001)
[17] Dörner, D.: The mathematics of emotions. In: Frank Detje, D.D., Schaub, H. (eds.) Proceedings of the Fifth International Conference on Cognitive Modeling, Bamberg, Germany, April 10-12, pp. 75–79 (2003)
[18] Damasio, A.: Descartes' Error: Emotion, Reason and the Human Brain. Gosset/Putnam Press, New York (1994)
[19] Ortony, A., Clore, G., Collins, A.: The cognitive structure of emotions. Cambridge University Press, Cambridge (1988)
[20] Lim, M.Y.: Emotions, Behaviour and Belief Regulation in An Intelligent Guide with Attitude. PhD thesis, School of Mathematical and Computer Sciences, Heriot-Watt University, Ediburgh, Edinburgh (2007)
[21] Ho, W.C., Dautenhahn, K., Nehaniv, C.L.: Computational memory architectures for autobiographic agents interacting in a complex virtual environment: A working model. Connection Science 20(1), 21–65 (2008)
[22] Lim, M.Y., Dias, J., Aylett, R., Paiva, A.: Improving adaptiveness in autonomous characters. In: Prendinger, H., Lester, J.C., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 348–355. Springer, Heidelberg (2008)

[23] Fails, J., Druin, A., Guha, M., Chipman, G., Simms, S., Churaman, W.: Child's play: A comparison of desktop and physical interactive environments. In: Conference on Interaction Design and Children, pp. 48–55 (2005)
[24] Dow, S., Mehta, M., Harmon, E., MacIntyre, B., Mateas, M.: Presence and engagement in an interactive drama. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 1487–1484 (2007)
[25] Paiva, A., Prada, R., Chaves, R., Vala, M., Bullock, A., Andersson, G., Höök, K.: Towards tangibility in gameplay: building a tangible affective interface for a computer game. In: ICMI 2003: Proceedings of the 5th international conference on Multimodal interfaces, pp. 60–67. ACM, New York (2003)
[26] Inkpen, K., Ho-Ching, W., Inkpen, K.H.C., Scott, S., Shoemaker, G.: This is fun! we're all best friends and we're all playing: supporting children's synchronous collaboration. In: Proceedings of the 1999 conference on Computer support for collaborative learning (1999)
[27] Mandryk, R., Inkpen, K., Bilezikjian, M., Klemmer, S., Landay, J.: Supporting children's collaboration across handheld computers. In: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, USA (2001)
[28] Leichtenstern, K., André, E., Vogt, T.: Role assignment via physical mobile interaction techniques in mobile multi-user applications for children. In: European Conference on Ambient Intelligence, Darmstadt, Germany (2007)
[29] Rehm, M., Bee, N., André, E.: Wave like an egyptian accelerometer based gesture recognition for culture specific interactions. In: HCI 2008 Culture, Creativity, Interaction (2008)
[30] Ang, S., Van Dyne, L., Koh, C., Ng, K.Y., Templer, K.J., Tay, C., Chandrasekar, N.A.: Cultural intelligence: Its measurement and effects on cultural judgment and decision making, cultural adaptation and task performance. Management and Organisation Review 3, 335–371 (2007)
[31] Stephan, W.G., Diaz-Loving, R., Duran, A.: Integrated threat theory and intercultural attitudes: Mexico and the united states. Journal of Cross-Cultural Psychology 31, 240–249 (2000)
[32] Wright, S.C., Aron, A., McLaughlin-Volpe, T., Ropp, S.A.: The extended contact effect: Knowledge of cross-group friendships and prejudice. Journal of Personality and Social Psychology 73, 73–90 (1997)

MusicCommentator: Generating Comments Synchronized with Musical Audio Signals by a Joint Probabilistic Model of Acoustic and Textual Features

Kazuyoshi Yoshii and Masataka Goto

National Institute of Advanced Industrial Science and Technology (AIST)
Central 2, 1-1-1 Umezono, Tsukuba, 305-8568 Ibaraki, Japan
{k.yoshii,m.goto}@aist.go.jp

Abstract. This paper presents a system called *MusicCommentator* that suggests possible comments on appropriate temporal positions in a musical audio clip. In an online video sharing service, many users can provide free-form text comments for temporal events occurring in clips not for entire clips. To emulate the commenting behavior of users, we propose a joint probabilistic model of audio signals and comments. The system trains the model by using existing clips and users' comments given to those clips. Given a new clip and some of its comments, the model is used to estimate what temporal positions could be commented on and what comments could be added to those positions. It then concatenates possible words by taking language constraints into account. Our experimental results showed that using existing comments in a new clip resulted in improved accuracy for generating suitable comments to it.

Keywords: Audio and language processing, user communication modeling, probabilistic music-comment association, comment generation.

1 Introduction

Commenting plays important roles in the entertainment culture of the consumer generated media (CGM). We can access many online content-sharing services such as YouTube (video), MySpace (music), and Flickr (photo) that enable users not only to present their original works but also to comment on works created by others. Users who view the same work can communicate with each other by commenting. For example, users provide tags [1], describe positive or negative reviews [2], and show their agreement or disagreement with specific comments. This kind of user communication has recently been facilitated.

Commenting can be viewed as collaborative creation [3] in an advanced form of user communication: pseudo-synchronized communication. We can see evidence of this in a video sharing service named *Nico Nico Douga* (whose name means "video making people smile" in Japanese),[1] where users can provide comments

[1] Nico Nico Douga reserves over 17 hundred million comments by ten million users. http://www.nicovideo.jp/, http://en.wikipedia.org/wiki/Nico_Nico_Douga.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 85–97, 2009.

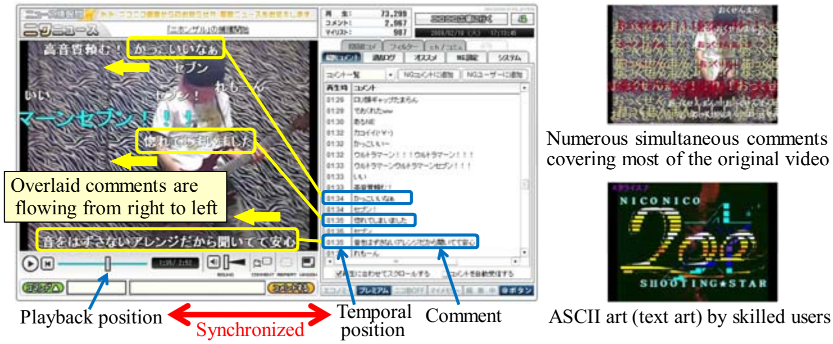


Fig. 1. Screenshots captured from a video sharing service named Nico Nico Douga

at arbitrary temporal positions in the video. A unique feature of this service is that recent comments of many users are overlaid directly onto the video and synchronized to a specific playback time as shown in Fig. 1. This gives users a sense of sharing the viewing experiences. That is, users feel as if they enjoyed the same video together in real time although their comments were provided at different dates and times in the real world. Some kinds of comments therefore can add remarkable and interesting effects to the original video. For instance, we often see *barrages*, where so many identical or similar comments collaboratively made by many users are piled up to a degree that the original video is almost completely hidden, and skilled users create cool drawings[2] by combining many comments (characters) provided at temporally and spatially different positions. Because commenting is much easier than creating new video clips from scratch or reediting existing ones. it can be an important popularized way of creation.

Novice users, however, sometimes feel anxiety when commenting, wondering *Is my comment suitable to the occasion?* Implicit rules seem to be shared among users who collaborate to provide comments, so novice users had better experience what kinds of comments are given by other users and what kinds of temporal events are annotated. Another issue is that video creators can hardly predict what comments will be given to their works, and the fear of being insulted often makes them hesitant about presenting their work to the public, especially the first time they try. Besides these practical issues, from an academic viewpoint, we are interested in the relationships between music and comments and investigate whether it is really possible to generate comments in a human-like fashion.

We therefore developed a system, called *MusicCommentator*, that can generate comments that are most likely to be provided at specific temporal positions in a music video clip. It can help novice users by suggesting comments suitable to the occasion and can help encourage video creators to present their work by letting them virtually experience having comments made about their work. In this study we deal with music, one of the audio parts within video clips, as the first step toward handling all the information in the clips.

[2] ASCII art or text art: http://en.wikipedia.org/wiki/ASCII_art.

The rest of this paper is organized as follows. First, Section 2 introduces related work. Then, Section 3 specifies the commenting problem and Section 4 explains how to build our system. Section 5 reports on our experiments. Finally, Section 6 summarizes the key findings of this paper.

2 Related Work

Several studies have been conducted to predict suitable words to a given musical piece by using the audio signal as input. These studies typically estimate how strongly each word is associated with a given piece. For example, Whitman and Rifkin [4] used a kernel method to predict words that will appear in music reviews. Turnbull *et al.* [5] attempted to associate audio content with semantically meaningful words by using a Gaussian mixture model (GMM) of acoustic features for each word. The output is obtained as sentences by filling slots with predicted words in sentence templates manually prepared beforehand. Bertin-Mahieux *et al.* [6] used an ensemble learning method called AdaBoost to predict social tags, which are free-form text labels at a song or artist level.

Our study differs from the previous studies in two ways. First, we deal with comments that are not given to an entire piece but provided at arbitrary temporal positions in it. It is thus necessary to determine what temporal positions can be annotated in a given piece. Second, we try to generate comments as natural-languages sentences. These goals make our attempt very challenging.

3 Problem Specification

The input data for the MusicCommentator task contains N audio clips (audio tracks of video clips) and their comments provided by users. Note that we focus on audio tracks in this paper even if we deal with video clips. Let n $(1 \leq n \leq N)$ be the index of an audio clip. This data is used to train a computational model of commenting. When the system is given a new audio clip (and some existing comments on it), the objective is to add reasonable comments at appropriate temporal positions by using the model. Audio clips are represented as acoustic features and comments are represented as textual features.

1. **Acoustic Features:** We use mel-frequency cepstrum coefficients (MFCCs) and their delta components because these features have been effectively used for characterization of detection of musical genres and moods [7]. Calculating MFCCs at each frame,[3] we can obtain a temporal sequence of feature vectors. Let $\boldsymbol{a}_t^{(n)}$ be a feature vector of frame t in clip n.
2. **Textual Features:** We define three kinds of textual features of comments.
   (a) **Bag-of-Words Features:** These features represent the content of provided comments. We split all Japanese free-form comments into words[4]

[3] A “frame” here is a short duration (256 ms) to be analyzed in an audio clip.

[4] Some words have single morphemes while others have two or more morphemes within them. The inflectional word “loved,” for example, consists of the base morpheme “love” and the inflectional morpheme “ed” (past tense).

with a Japanese morphological analyzer called Mecab [8]. Because different words including the same base morpheme are semantically identical, we do not distinguish them. Removing auxiliary words and extracting significant words whose numbers of occurrences are higher than a threshold, we get a vocabulary consisting of V words. Then, we count how many times each word occurs. For example, if a frame contains three comments, "I love it," "It is loved," and "Love song," the average number of occurrences of the verb "love" is 0.66. Let $\boldsymbol{w}_t^{(n)} = \{w_{t,1}^{(n)}, \cdots, w_{t,V}^{(n)}\}$ be a bag-of-words vector, where $w_{t,v}^{(n)}$ $(1 \leq v \leq V)$ represents the number of occurrences of word v per comment at frame t in clip n.

(b) **Comment Density:** This indicates the number of comments in each frame. Note that feature values in each clip are normalized with respect to its length and the number of comments. This feature is used to learn what temporal positions should be annotated in a target clip. Let $d_t^{(n)}$ be a feature value of frame t in clip n.

(c) **Average Length of Comments:** This indicates the average number of words in a single comment and is used to learn how long comments could be generated. In the above example with "love", the value of this feature is 2.66 ((3 + 3 + 2) / 3). Let $l_t^{(n)}$ be an average length of comments of frame t in clip n.

A set of these features is given by $\boldsymbol{o}_t^{(n)} = \{\boldsymbol{a}_t^{(n)}, \boldsymbol{w}_t^{(n)}, d_t^{(n)}, l_t^{(n)}\}$. When clip n contains T_n frames, the observable features $\boldsymbol{O}^{(n)}$ and $\boldsymbol{O}$ are given by $\boldsymbol{O}^{(n)} = \{\boldsymbol{o}_1^{(n)}, \cdots, \boldsymbol{o}_{T_n}^{(n)}\}$ and $\boldsymbol{O} = \{\boldsymbol{O}^{(1)}, \cdots, \boldsymbol{O}^{(N)}\}$.

4 MusicCommentator

MusicCommentator takes a constructive approach that tries to clarify the cognitive mechanism of humans by building and examining a computational model emulating it. As shown in Fig. 2, the system comprises a *learning* phase in which the system tries to acquire a sense of commenting (i.e., build a computational model of what comments are suitable to specific acoustic features) by experiencing many comments provided by users, and a *commenting* phase in which the model is used to generate comments suitable to the occasion. We will discuss how to design the model and then explain the two phases.

4.1 Model Formulation

Considering the characteristics of target data, we think that a reasonable model should meet the following three requirements:

1. **Joint Modeling of Acoustic and Textual Features:** When users want to produce new comments, they seem to simultaneously take into account the content of musical audio signals and the content of comments provided by other users. This suggests that well-balanced integration of them will enable the model to yield reasonable comments.

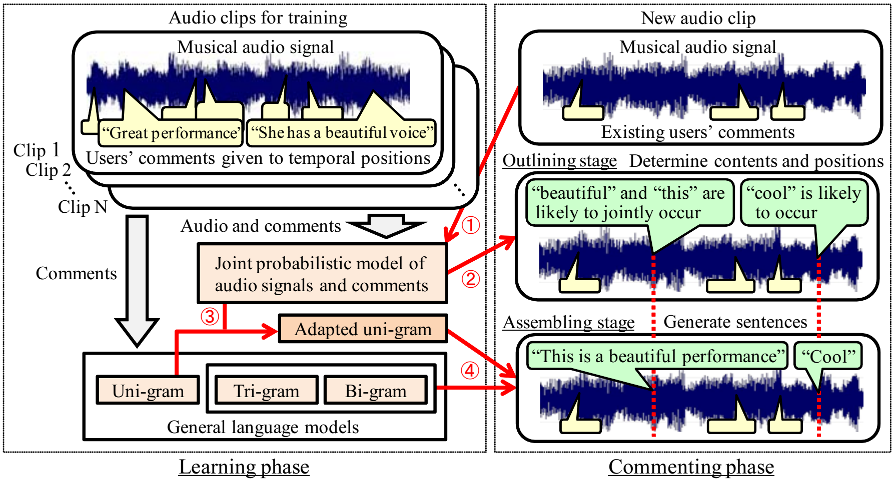


Fig. 2. Overview of MusicCommentator

2. **Temporal Modeling of Acoustic and Textual Features:** Because music is one of temporal medium, its temporal characteristics such as mood transitions should be captured by using a temporal model. Similarly, we focus on topic transitions in comments that are synchronized with music.
3. **Cross-modal Feature Binding through Temporal Contexts:** We can assume that a single latent *state*, which can be conceptually interpreted as a mood or topic, is shared behind the observable features of audio signals and comments at each frame.

To meet these requirements, we propose a joint probabilistic model of multimodal features by extending a standard hidden Markov model (HMM), as shown in Fig. 3. Let K be the number of latent states and let $\boldsymbol{z}_t^{(n)} = \{z_{t,1}^{(n)}, \cdots, z_{t,K}^{(n)}\}$ be a state representation at frame t in clip n, where $z_{t,k'}^{(n)} = 1$ and $\{z_{t,k}^{(n)} = 0 | k \neq k'\}$ if the model stays at state k' $(1 \leq k' \leq K)$. We define latent state sequences $\boldsymbol{Z}^{(n)}$ and $\boldsymbol{Z}$ as $\boldsymbol{Z}^{(n)} = \{\boldsymbol{z}_1^{(n)}, \cdots, \boldsymbol{z}_{T_n}^{(n)}\}$ and $\boldsymbol{Z} = \{\boldsymbol{Z}^{(1)}, \cdots, \boldsymbol{Z}^{(N)}\}$.

Our HMM can, like standard HMMs, be characterized by a set $\boldsymbol{\theta}$ of three kinds of parameters $\{\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{\phi}\}$. $\boldsymbol{\pi}$ is a set of initial probabilities $\{\pi_1, \cdots, \pi_K\}$, where $\pi_k \equiv p(z_{1,k}^{(\cdot)} = 1)$. $\boldsymbol{A}$ is a transition matrix $\{A_{jk} | 1 \leq j, k \leq K\}$, where $A_{jk} \equiv p(z_{t,k}^{(\cdot)} = 1 | z_{t-1,j}^{(\cdot)} = 1)$. $\boldsymbol{\phi}$ is a set of parameters of output distributions that calculate the likelihoods of observable features.

Acoustic and textual features at a frame are associated with the same state. Let b_k be a joint output distribution of state k. This calculates the likelihood of $\boldsymbol{o}_t^{(n)}$, which is given by $b_k(\boldsymbol{o}_t^{(n)})$. This indicates how likely the four kinds of features $\{\boldsymbol{a}_t^{(n)}, \boldsymbol{w}_t^{(n)}, d_t^{(n)}, l_t^{(n)}\}$ jointly occur from state k. We assume that $b_k(\boldsymbol{o}_t^{(n)})$ can be decomposed into the following four likelihoods:

$$b_k(\boldsymbol{o}_t^{(n)}) = b_{a,k}(\boldsymbol{a}_t^{(n)})\ b_{w,k}(\boldsymbol{w}_t^{(n)})\ b_{d,k}(d_t^{(n)})\ b_{l,k}(l_t^{(n)}), \tag{1}$$

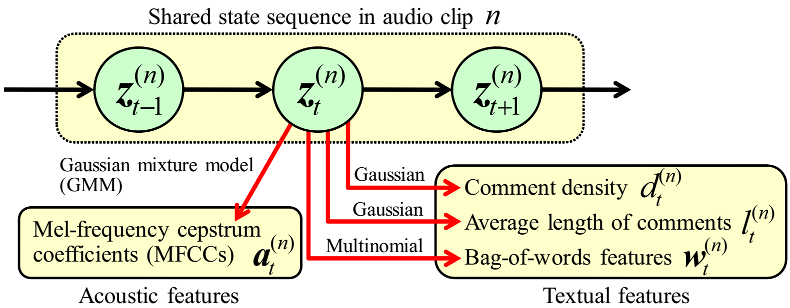


Fig. 3. Overview of our ergodic hidden Markov model

where $b_{a,k}$, $b_{w,k}$, $k_{d,k}$, and $b_{l,k}$ are designed as follows:

1. **Output Distributions of Acoustic Features:** $b_{a,k}$ is a Gaussian mixture model (GMM) of state k as in typical HMMs for speech recognition. Let M be the number of mixtures and let $g_{a,k,m}$, $\boldsymbol{\mu}_{a,k,m}$, and $\boldsymbol{\Sigma}_{a,k,m}$ be the weight, mean, and covariance of the m-th Gaussian in the GMM of state k.
2. **Output Distributions of Textual Features:** $b_{w,k}$ is a multinomial distribution for bag-of-words features. Its parameters are given by $\boldsymbol{p}_k = \{p_{k,1}, \cdots, p_{k,V}\}$. $b_{d,k}$ is a standard Gaussian representing the distribution of comment densities. Let $\mu_{d,k}$ and $\Sigma_{d,k}$ be the mean and variance of the Gaussian associated with state k. Similarly, $b_{l,k}$ is also a Gaussian for comment lengths, and its mean and variance are given by $\mu_{l,k}$ and $\Sigma_{l,k}$.

Let $\boldsymbol{\phi}_k$ be the set of parameters of output distributions of state k, given by $\boldsymbol{\phi}_k = \{\{g_{a,k,m}, \boldsymbol{\mu}_{a,k,m}, \boldsymbol{\Sigma}_{a,k,m} | 1 \leq m \leq M\}, \boldsymbol{p}_k, \mu_{d,k}, \Sigma_{d,k}, \mu_{l,k}, \Sigma_{l,k}\}$. In total, a set of all parameters of output distributions $\boldsymbol{\phi}$ is given by $\boldsymbol{\phi} = \{\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_K\}$.

Our model is an ergodic HMM, which allows any state transition at any time, because we cannot identify correct sequences of states in training data. In speech recognition, manual transcriptions of speech signals (i.e., phoneme sequences) can be directly transformed into state sequences because each phoneme is defined as a combination of several states. Left-to-Right HMMs, where state transitions are limited to match the transcriptions, are therefore commonly used in speech recognition. In contrast, we use the HMM in an unsupervised fashion.

4.2 Learning Phase

This section explains how to estimate the unknown parameters $\boldsymbol{\theta} = \{\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{\phi}\}$. Let $p(\boldsymbol{O}|\boldsymbol{\theta})$ be the likelihood of observable variables $\boldsymbol{O}$. Instead of directly maximizing the *incomplete* likelihood $p(\boldsymbol{O}|\boldsymbol{\theta})$, we try to maximize the expected *complete* likelihood of observable variables $\boldsymbol{O}$ and latent variables $\boldsymbol{Z}$ by using the Expectation-Maximization (EM) algorithm [9]. The complete likelihood is

$$p(\boldsymbol{O}, \boldsymbol{Z}|\boldsymbol{\theta}) = \prod_{n=1}^{N} p(\boldsymbol{z}_1^{(n)}|\boldsymbol{\pi}) \left[\prod_{t=2}^{T_n} p(\boldsymbol{z}_t^{(n)}|\boldsymbol{z}_{t-1}^{(n)})\right] \prod_{t=1}^{T_n} p(\boldsymbol{o}_t^{(n)}|\boldsymbol{z}_t^{(n)}), \tag{2}$$

where $p(\boldsymbol{z}_1^{(n)}|\boldsymbol{\pi})$ is given by $\prod_{k=1}^{K} \pi_k^{z_{1,k}^{(n)}}$, which is an initial probability that clip n starts at a state specified by $\boldsymbol{z}_1^{(n)}$. We then define the Q function as follows:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{old}) = \sum_{\boldsymbol{Z}} p(\boldsymbol{Z}|\boldsymbol{O}, \boldsymbol{\theta}_{old}) \log p(\boldsymbol{O}, \boldsymbol{Z}|\boldsymbol{\theta}), \tag{3}$$

where $\boldsymbol{\theta}_{old}$ is a set of the current parameters and $p(\boldsymbol{Z}|\boldsymbol{O}, \boldsymbol{\theta}_{old})$ is a posterior probability of latent variables $\boldsymbol{Z}$. $Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{old})$ indicates the *expected* complete log-likelihood of all variables $\boldsymbol{O}$ and $\boldsymbol{Z}$ when we regard $\boldsymbol{\theta}$ as a variable of the function. Thus, the optimized parameters are obtained by maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{old})$ and are then set to $\boldsymbol{\theta}_{old}$ next time. This is iterated until $Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{old})$ converges.

In the E-step of the EM algorithm, the objective is to calculate the posterior distribution $p(\boldsymbol{Z}|\boldsymbol{O}, \boldsymbol{\theta}_{old})$. For convenience, we define some new symbols as:

$$\gamma(\boldsymbol{z}_t^{(n)}) \equiv p(\boldsymbol{z}_t^{(n)}|\boldsymbol{O}, \boldsymbol{\theta}_{old}),\ \ \xi(\boldsymbol{z}_{t-1}^{(n)}, \boldsymbol{z}_t^{(n)}) \equiv p(\boldsymbol{z}_{t-1}^{(n)}, \boldsymbol{z}_t^{(n)}|\boldsymbol{O}, \boldsymbol{\theta}_{old}), \tag{4}$$

$$\gamma(\boldsymbol{y}_{t,k}^{(n)}) \equiv p(\boldsymbol{y}_{t,k}^{(n)}|\boldsymbol{O}, \boldsymbol{\theta}_{old}) = p(\boldsymbol{y}_{t,k}^{(n)}|\boldsymbol{z}_t^{(n)})\gamma(\boldsymbol{z}_t^{(n)}), \tag{5}$$

where $\gamma(\boldsymbol{z}_t^{(n)})$ is a posterior distribution of latent variable $\boldsymbol{z}_t^{(n)}$. $\xi(\boldsymbol{z}_{t-1}^{(n)}, \boldsymbol{z}_t^{(n)})$ is a joint posterior distribution of adjacent latent variables $\boldsymbol{z}_{t-1}^{(n)}$ and $\boldsymbol{z}_t^{(n)}$. For each t, $\gamma(\boldsymbol{z}_t^{(n)})$ consists of K probabilities that sum up to unity. $\xi(\boldsymbol{z}_{t-1}^{(n)}, \boldsymbol{z}_t^{(n)})$ is expressed as a $K \times K$ probability matrix whose elements sum up to unity. Let $\gamma(z_{t,k}^{(n)})$ be the conditional probability of $z_{t,k}^{(n)} = 1$ and let $\xi(z_{t-1,j}^{(n)}, z_{t,k}^{(n)})$ be that of $z_{t-1,j}^{(n)} = z_{t,k}^{(n)} = 1$, given $\boldsymbol{O}$ and $\boldsymbol{\theta}_{old}$. These probabilities can be efficiently calculated by using the forward-backward algorithm [11]. $\boldsymbol{y}_{t,k}^{(n)}$ is a vectorial variable, $\{y_{t,k,1}^{(n)}, \cdots, y_{t,k,M}^{(n)}\}$. This shows which Gaussian is responsible for generating $\boldsymbol{a}_t^{(n)}$ among M Gaussians in GMM $b_{a,k}$, where $y_{t,k,m'}^{(n)} = 1$ and $\{y_{t,k,m}^{(n)} = 0|m \neq m'\}$ when the m'-th Gaussian is responsible. $p(\boldsymbol{y}_{t,k}^{(n)}|\boldsymbol{z}_t^{(n)})$ is expressed as a $K \times M$ probability matrix that reserves the responsibilities of KM Gaussians of GMMs $\{b_{a,1}, \cdots, b_{a,K}\}$ for observation $\boldsymbol{a}_t^{(n)}$.

In the M-step, we try to maximize the Q function $Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{old})$. Substituting Eqn. (2) for Eqn. (3), we get

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}_{old}) = \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{1,k}^{(n)}) \log \pi_k + \sum_{n=1}^{N}\sum_{t=2}^{T_n}\sum_{j=1}^{K}\sum_{k=1}^{K} \xi(z_{t-1,j}^{(n)}, z_{t,k}^{(n)}) \log A_{jk} \\ + \sum_{n=1}^{N}\sum_{t=1}^{T_n}\sum_{k=1}^{K} \gamma(z_{t,k}^{(n)}) \log p(\boldsymbol{o}_t^{(n)}|\boldsymbol{\phi}_k), \end{aligned} \tag{6}$$

where the last term can be decomposed into four terms as $\log p(\boldsymbol{o}_t^{(n)}|\boldsymbol{\phi}_k) = \log b_{a,k}(\boldsymbol{a}_t^{(n)}) + \log b_{w,k}(\boldsymbol{w}_t^{(n)}) + \log b_{d,k}(d_t^{(n)}) + \log b_{l,k}(l_t^{(n)})$ We can thus independently update the parameters of four kinds of distributions (GMM $b_{a,k}$, multinomial distribution $b_{w,k}$, and two Gaussians $b_{d,k}$ and $b_{l,k}$) by using the Lagrange multiplier method. We get the updating formula as follows:

$$\pi_k = \frac{\sum_{n=1}^{N} \gamma(z_{1,k}^{(n)})}{\sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{1,k}^{(n)})}, \ A_{jk} = \frac{\sum_{n=1}^{N} \sum_{t=2}^{T_n} \xi(z_{t-1,j}^{(n)}, z_{t,k}^{(n)})}{\sum_{n=1}^{N} \sum_{l=1}^{K} \sum_{t=2}^{T_n} \xi(z_{t-1,j}^{(n)}, z_{t,l}^{(n)})},$$

$$g_{a,k,m} = \frac{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \gamma(y_{t,k,m}^{(n)})}{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \sum_{m=1}^{M} \gamma(y_{t,k,m}^{(n)})}, \ \boldsymbol{\mu}_{a,k,m} = \frac{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \gamma(y_{t,k,m}^{(n)}) \boldsymbol{a}_t^{(n)}}{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \gamma(y_{t,k,m}^{(n)})},$$

$$\boldsymbol{\Sigma}_{a,k,m} = \frac{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \gamma(y_{t,k,m}^{(n)}) (\boldsymbol{a}_t^{(n)} - \boldsymbol{\mu}_{a,k,m})^2}{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \gamma(y_{t,k,m}^{(n)})}, \ \boldsymbol{p}_k = \frac{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \gamma(z_{t,k}^{(n)}) \boldsymbol{w}_t^{(n)}}{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \gamma(z_{t,k}^{(n)})},$$

$$\mu_{d,k} = \frac{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \gamma(z_{t,k}^{(n)}) d_t^{(n)}}{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \gamma(z_{t,k}^{(n)})}, \ \Sigma_{d,k} = \frac{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \gamma(z_{t,k}^{(n)}) (d_t^{(n)} - \mu_{d,k})^2}{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \gamma(z_{t,k}^{(n)})},$$

$$\mu_{l,k} = \frac{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \gamma(z_{t,k}^{(n)}) l_t^{(n)}}{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \gamma(z_{t,k}^{(n)})}, \ \Sigma_{l,k} = \frac{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \gamma(z_{t,k}^{(n)}) (l_t^{(n)} - \mu_{l,k})^2}{\sum_{n=1}^{N} \sum_{t=1}^{T_n} \gamma(z_{t,k}^{(n)})}. \quad (7)$$

4.3 Commenting Phase

This section explains how to provide comments suitable to a target audio clip. Like the training data, the audio signal and provided comments are characterized by a sequence of acoustic features, $\boldsymbol{a}' = \{\boldsymbol{a}'_1, \cdots, \boldsymbol{a}'_{T'}\}$, three sequences of textual features, $\boldsymbol{w}' = \{\boldsymbol{w}'_1, \cdots, \boldsymbol{w}'_{T'}\}$, $\boldsymbol{d}' = \{d'_1, \cdots, d'_{T'}\}$, and $\boldsymbol{l}' = \{l'_1, \cdots, l'_{T'}\}$, where T' is the number of frames. This phase consists of an *outlining* stage and an *assembling* stage. The latter estimates how many comments and what content should be provided at each frame. The former concatenates a suitable number of words in an appropriate order by taking language constraints into account.

Outlining Stage. We first determine a most likely sequence of latent states in the target, $\boldsymbol{z}' = \{\boldsymbol{z}'_1, \cdots, \boldsymbol{z}'_{T'}\}$, with the Viterbi algorithm [10]. When $z'_{t,k}$ is 1 at frame t ($1 \leq t \leq T'$), the most likely density there, $\hat{d}_t$, is given by the mode (most likely observation) of the Gaussian $b_{d,k}$, i.e., mean $\mu_{d,k}$. From the density distribution over the entire clip, we can determine how many comments should be provided in each frame. Similarly, when $z'_{t,k}$ is 1, we can get most likely bag-of-words features (occurrence probabilities of significant words) $\hat{\boldsymbol{w}}_{\boldsymbol{t}}$ to be $\boldsymbol{p}_k$.

We here cannot generate sentences that are appropriate as natural language, i.e., reasonable sequences of words, because bag-of-words features only outlines the content of comments. Therefore, we should solve the following problems:

1. We do not have occurrence probabilities of non-significant words such as conjunctions and auxiliary verbs, which are indispensable for natural language.
2. We do not have individual occurrence probabilities of inflectional words that have the same base morpheme within them (see 2a in Section 3).

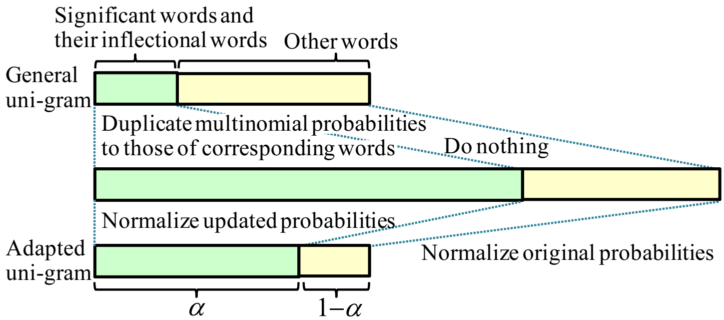


Fig. 4. Adaptation of general uni-gram to multinomial distribution

3. We cannot determine an appropriate order of words because the current model does not take into account sequential relations between words.

For example, suppose that two words "it" and "love" are highly likely to occur and the comment length is likely to be three. We cannot synthesize a comment like "It is loved" or "I love it" because the probabilities of "is" and "loved" are not given and we therefore do not know which sentence is more appropriate.

Assembling Stage. To solve the three problems described above, we propose a comment generation method based on adaptation of general language models (uni-, bi-, and tri-grams) that are learned from numerous comments of all clips in the training data. Unlike what we do in the learning phase, we distinguish between different words that have the same base morpheme (e.g., we distinguish "took" from "taken"'). The uni-gram can be used for solving the first and second problems, and the bi- and tri- grams contribute to solving the third one.

Fig. 4 shows a sketch of how the probabilities $\hat{\boldsymbol{w}}_{\boldsymbol{t}}$ of significant words at frame t are incorporated into the general uni-gram, which includes all the words appearing in the training data. We duplicate the probability of each significant word to those of its inflectional words that have different surface expressions. For example, the probabilities of words "took" and "taken" are set to be the same as that of word "take." Then, because the sum the probabilities of significant words and their inflectional ones is greater than 1, the probabilities are scaled so that their sum is α, which is a control parameter given in advance. On the other hand, the probabilities of other words containing non-significant words and their inflectional ones are scaled so that their sum is $1-\alpha$. As a result, we now get the adapted uni-gram (a set of occurrence probabilities of *all* words).

When $z'_{t,k}$ is 1 at frame t, we assume that a most likely comment (word sequence) $\hat{\boldsymbol{c}}_t$ and a most likely comment length (number of words) $\hat{l}_t$ should be determined according to the following generative model:

$$\{\hat{\boldsymbol{c}}_t, \hat{l}_t\} = \underset{\boldsymbol{c},l}{\operatorname{argmax}}\, p(\boldsymbol{c}, l; \boldsymbol{\theta}_k) = \underset{\boldsymbol{c},l}{\operatorname{argmax}}\, p(\boldsymbol{c}|l; \boldsymbol{\theta}_k) p(l; \boldsymbol{\theta}_k), \tag{8}$$

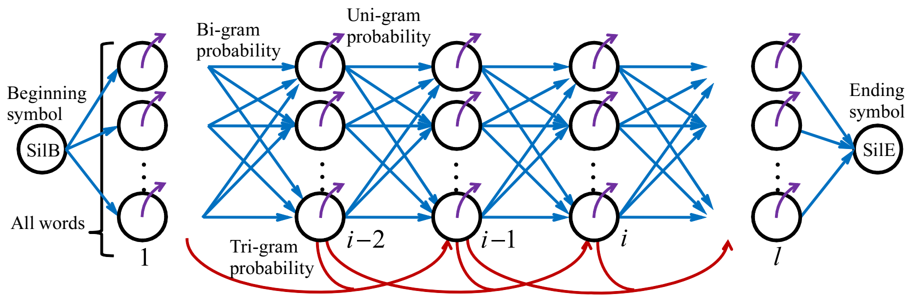


Fig. 5. Probability calculation on word trellis

where $p(l;\boldsymbol{\theta}_k)$ is a likelihood that a comment generated from state k consists of l words. Its value is calculated according to Gaussian $b_{l,k}$ and $p(\boldsymbol{c}|l;\boldsymbol{\theta}_k)$ is a conditional probability that comment $\boldsymbol{c}$ is generated when its length is given by l. Note that for readability we hereafter omit the estimated parameter $\boldsymbol{\theta}_k$. To get $\hat{\boldsymbol{c}}$, we have only to calculate $\mathrm{argmax}_{\boldsymbol{c}}\, p(\boldsymbol{c}|l)$ for each length.

To estimate $\mathrm{argmax}_{\boldsymbol{c}}\, p(\boldsymbol{c}|l)$, we propose a method that can find a most likely path of words on a trellis including all words by using the Viterbi algorithm [10]. As shown in Fig. 5, each node corresponds to a specific word and the observation probabilities of words in each column are given by the adapted uni-gram. Transition probabilities between nodes are determined as bi- and tri-grams. We let SilB and SilE be special symbols (silent words) that indicate the beginning and ending of comments. The likelihood of comment $\boldsymbol{c}$ is given by

$$p(\boldsymbol{c}|l) = p(w_1|\mathrm{SilB}) \left(\prod_{i=2}^{l} p(w_i|w_{i-2}, w_{i-1}) \right) p(\mathrm{SilE}|w_{l-1}, w_l), \tag{9}$$

where w_i is the i-th word in comment $\boldsymbol{c}$ and w_0 is SilB. $p(w_i|w_{i-2}, w_{i-1})$ is an *adapted* trigram probability, which is calculated with linear interpolation of the general tri- and bi- grams p_t, p_b and the adapted uni-grams p'_u as follows: $p(w_i|w_{i-2}, w_{i-1}) \leftarrow \beta_t p_t(w_i|w_{i-2}, w_{i-1}) + \beta_b p_b(w_i|w_{i-1}) + \beta_u p'_u(w_i)$, where β_t, β_b, and β_u are weighting factors of the tri-, bi-, and uni-grams.

5 Evaluation

We experimentally evaluated how accurately the system predicted comments that are freely provided on arbitrary temporal positions by users.

5.1 Conditions

The audio clips (tracks) we used were included in the video clips taken from the music category of the video sharing service *Nico Nico Douga*. Specifically, we focus on music performances whose titles included *"Ensoushitemita"* ("We/I played

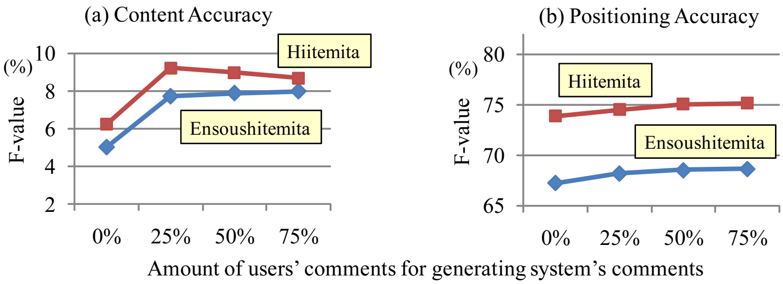


Fig. 6. Results of experimental comment generation

something, not limited to musical instruments, e.g., music box and wooden gong"), or *"Hiitemita"* ("We/I played piano or stringed instruments, e.g., violin and guitar"). Some of them were performed by multiple people, e.g., sessions, bands, or ensembles. There are many popular clips that follow these conventional naming rules in the title. We collected the most popular 100 *"Ensoushitemita"*-category clips that are shorter than 10 minutes according the the number of comments that roughly reflects its popularity. Then, the first 1100 comments, which were available as many as possible in all the clips, were extracted. Note that the first 1100 comments do not mean the 1100 comments taken from the beginning within each clip, but mean the 1100 comments taken from the beginning of its submission to the video sharing service. As for the *"Hiitemita"* category having more comments, we were able to extract 2400 comments from each of 100 clips. The control parameters were set as $V = 2082, 2278$, $K = 200$, $M = 8$, $\alpha = 0.9$, and $\beta_t = \beta_b = \beta_u = 1.0$ by trial and error.

The experiments were conducted in the way of 4-fold cross fold validation. First, all audio clips with provided comments were randomly divided into four groups. Three groups were used as a training set in the learning phase and the other group was used as a test set in the commenting phase. Switching the choice of test set, we conducted four trials. 4-fold cross fold validation was furthermore performed in each trial by dividing the provided comments of each test clip into four groups. To estimate a most likely sequence of states in a given clip, the system used either no comments (i.e., only acoustic features) or one, two, or three groups (i.e., 0%, 25%, 50%, or 75%) of comments on the clip. That is, we tested four settings. The remaining comments was used as ground truth.

To evaluate the results, we calculated the word-based F-value, which is given by $F = \frac{2PR}{P+R}$, where P and R are the precision and recall rates. We focused on each word of the system-generated comments. A word in a system's comment was considered reasonable if it appeared in users' comments annotated in the neighborhood of the system's comment. The error tolerance was set to 5 seconds.

$$P = \frac{\#\text{appropriate words}}{\#\text{words of system's comments}}, \quad R = \frac{\#\text{appropriate words}}{\#\text{words of users' comments}}, \tag{10}$$

5.2 Results

As shown in Fig. 6(a), the F-values could be improved even if only 25% of users' comments of a target clip was available for adding new comments. Although the F-values reached at most 10%, we think these results were promisingly reasonable because it is impossible to completely predict what comments are provided by users at a word level even for humans. Note that when we evaluated only the temporal positions and lengths of generated comments (i.e., allowed errors in word selection), the F-values were around 70%, as shown in Fig. 6(b). One may say that it is enough to list most likely words as a rough suggestion. However, we believe that sentences of natural language are much better in terms of readability although they are often grammatically or semantically strange because n-grams cannot all inter-word dependencies contained in a sentence.

The F-values were not furthermore improved when we increased the amount of users' comments over 25% for adding new comments. This indicates that the current system cannot deal with widely diverse comments. That is, the model cannot create various comments that are essentially different from each other in their meanings once a specific state is determined for given acoustic and textual features. Comments freely provided by humans without constraints are widely diverse. A major reason that the F-values for the *"Ensoushitemita"* category were lower than those for the *"Hiitemita"* category could be the wider diversity of the comments on former clips. The title *"Hiitemita"* means limited kinds of instruments such as piano and guitar were used in the video.

We also found that the current system is not always useful because general comments like "it is very cool" and "great" tend to be generated. The F-values of frequently used positive words such as "cool" or "great" were around 40%. This was closely related to the limitation of the statistical approach. If we can use a huge amount of users' comments for training the HMM (we actually used over 100,000 comments), the probabilistic model tries to capture universal characteristics of the data. However, it is not appropriate to spoil the diversity of humans' comments for our task. We should tackle this problem in the future.

6 Conclusion

We presented MusicCommentator that generates comments (short sentences of natural language) and provide them at appropriate temporal positions. The system is based on a multi-modal HMM that associates acoustic features with textual ones through latent sequences of states. These sequences correspond to temporal transitions of both musical moods and comment topics. To estimate the parameters of the HMM, we used a likelihood maximization method so that many examples of how users have provided comments can be well explained with the model. Given a new audio clip, the system concatenates suitable words in an appropriate order by using general language models.

The experimental results were promising but revealed that we are still far from the ultimate goal of building a computer that can express the impressions of video clips as natural language as humans do. Because commenting is one of the most sophisticated cognitive functions of humans, it would be hard to precisely emulate even if we use the state-of-the-art techniques of machine learning. We think, however, that our study is an important first challenge. We plan to improve MusicCommentator by incorporating advanced methods of recognizing musical content such as rhythm and melody. This kind of multi-aspect modeling could help the system generate comments that are more appropriate and diverse.

Acknowledgement. This study was partially supported by CREST, JST.

References

1. Ames, M., Naaman, M.: Why We Tag: Motivations for Annotation in Mobile and Online Media. In: ACM CHI, pp. 971–980 (2007)
2. Nakamura, S., Shimizu, M., Tanaka, K.: Can Social Annotation Support Users in Evaluating the Trustworthiness of Video Clips? In: ACM WICOW (2008)
3. Hamasaki, M., Takeda, H., Nishimura, T.: Network Analysis of Massively Collaborative Creation of Multimedia Contents –Case Study of Hatsune Miku videos on Nico Nico Douga–. In: uxTV, pp. 165–168 (2008)
4. Whitman, B., Rifkin, R.: Musical Query-by-Description as a Multiclass Learning Problem. In: IEEE MMSP 2002, pp. 153–156 (2002)
5. Turnbull, D., Barrington, L., Torres, D., Lanckriet, G.: Semantic Annotation and Retrieval of Music and Sound Effects. IEEE Trans. on ASLP 16(2), 467–476 (2008)
6. Bertin-Mahieux, T., Eck, D., Maillet, F., Lamere, P.: Autotagger: A Model for Predicting Social Tags from Acoustic Features on Large Music Databases. J. of New Music Research (JNMR) 37(2), 115–135 (2008)
7. Tzanetakis, G., Cook, P.: Musical Genre Classification of Audio Signals. IEEE Trans. on SAP 10(5), 293–302 (2002)
8. Kudo, T., Yamamoto, T., Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis. In: EMNLP (2004)
9. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm. J. R. Stat. Soc (B) 39(1), 1–38 (1977)
10. Forney, G.D.: The Viterbi Algorithm. IEEE 61(3), 268–278 (1973)
11. Baum, L.E., Petrie, T., Soules, G., Weiss, N.: A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. Annals of Mathematical Statistics 41(1), 164–171 (1970)

MiniDiver: A Novel Mobile Media Playback Interface for Rich Video Content on an iPhone™

Gregor Miller[1], Sidney Fels[1,2], Matthias Finke[2], Will Motz[1], Walker Eagleston[1], and Chris Eagleston[1]

[1] Dept. of Electrical & Computer Engineering, University of British Columbia, Vancouver, BC, Canada
[2] Media and Graphics Interdisciplinary Centre, University of British Columbia, Vancouver, BC, Canada

Abstract. We describe our new mobile media content browser called a *MiniDiver*. MiniDiving considers media browsing as a personal experience that is viewed, personalized, saved, shared and annotated. When placed on a mobile platform, such as the iPhone™, consideration of the particular interface elements lead to new ways to experience media content. The MiniDiver interface elements currently supports multi-camera selection, video hyperlinks, history mechanisms and semantic and episodic video search. We compare performance of the MiniDiver on different media streams to illustrate its feasibility.

1 Introduction

Currently, on a mobile platform, video is normally watched with a video browser that still uses a tape playing metaphor. That is, it has transport controls for *play*, *rewind*, *forward*, *skip forward* and *backward*. However, when video contains metadata, such as hyperlinks, semantic content (i.e., keywords) and multiple camera angles, the interface does not allow a simple way to navigate through this complex video space. Thus, for example, when browsing YouTube™ it is easy to get lost in a multitude of video sources. As well, there is no obvious way to record history or share partial content or provide annotation. Likewise, mechanisms to hyperlink content or do episodic or semantic based searches are limited. We have been exploring new interaction paradigms for video browsing, navigation and annotation as part of our *MyView* research to provide personalized video experiences using a novel video browsing MyView client called a *Diver*.

The concept behind the MyView research program is that in the future, tracking technology and pervasive video/audio capture technology will be used to automatically tag video content making for rich, complex video data space that can be viewed differently by each person. In this paradigm, the notion of the video clip becomes less clear since it can be at the granularity of a single person in a single instant. For example, at an indoor Olympic event, such as ice hockey, we can create personal views of the game while it is being played as well as

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 98–109, 2009.

Fig. 1. Interface of Video Player with Heads Up Display (HUD) and Toolbar

after the game that can be delivered over wireless protocols to cell phones or other devices. Audio and video data may be captured by multiple sensors and streamed to a multimedia server continuously. Hockey players may be tracked using video and/or other mechanisms to detect when they are in view of each video camera providing tracking and view orientation meta-data. Likewise, additional semantic data, such as keywords, may be provided by broadcasters and other viewers. With this meta-data combined with the multiple video sources, a valuable, personalized, fun memory of an event can be viewed and shared. The Diver provides mechanisms to view this rich video space, though, is currently intended to be embedded in web-browsers and WIMP interfaces.

As part of MyView, we also address the emerging trend that video entertainment on a mobile platform is quickly becoming typical. Thus we also have constructed a version of the Diver for a mobile environment; specifically, an iPhone™. We call this a *MiniDiver* and it takes advantage of the specifics of limited screen real-estate and particular input mechanisms found on mobile phones. Figure 1 shows an example of the iPhone™ interface for the MiniDiver. In this paper, we focus on four main issues that arise when rich video content is viewed on the cell phone. Specifically, we have investigated interface mechanisms in the MiniDiver to allows users to: 1) select multiple camera views, 2) navigate hyperlinked video content, 3) save and retrieve complex MiniDiving history, and 4) use both episodic and semantic mechanisms to access video meta-data.

Each of these four items are emerging as key components of the video experience. For example, at a sporting event, such as ice hockey, there may be multiple cameras focused on the action synchronized in time. Likewise, each player's location may be tracked and appropriate hyperlinks to new video or other web content be integrated with the video experience. As users Dive through the hockey game, their history becomes complex as they follow hyperlinks and change camera angles. In this situation, going "back" isn't obvious for what should happen. Finally, the meta-data associated with the data allows for automatic assembly of personalized video based on combinations of episodic (i.e, image/video based) or semantic (i.e. keywords) mechanisms. These are integrated into the MiniDiver as we discuss below.

In Section 2 we cover some of the related work that has investigated rich video content mechanism and means to access it by users. Section 3 provides a description of the main types of interface paradigms we have created for the video experience. All our examples use ice hockey as our event that we are MiniDiving. We have used multiple cameras to capture footage of an ice hockey match and have created meta-data tags associated with the players from all camera angles for our test data. We cover some of the performance issues in Section 4 and conclude with a discussion of the emerging complexity of navigating rich video data spaces and the interaction design approaches needed to deal with them.

2 Related Work

For many years traditional TV stations utilize multi-camera views to enhance the user's experience especially when broadcasting sport events. Relaying on a broadcast channel viewers are dependent on directors who select camera angles to provide a "best view" for an entire audience. Taking the "best view" experience to the next level implies viewers can choose for themselves the "best view" based on their personal preference. This requires new interface concepts for video players that enable viewers to switch between available camera views. Lou et al. [1] developed a video player for their multi-view video system which includes a slider within the interface to allow viewers to switch seamlessly among views of the current video presentation. Navigational aids within the video space are not supported, which makes it especially difficult for longer video sequences including multiple views.

In computer vision research multi-camera views are widely used, e.g. for object tracking [2] or view synthesis[3]. Sport video analysis especially takes advantage of object tracking algorithms that employ multiple camera sources in order to extract context sensitive metadata e.g. player location [4]. Such location data can be used to create object-based video annotations [5], also called hypervideo links[6]. Similar to a hyperlink in a web page viewers can interact with a hypervideo link to access additional information. Hypervideo links follow the associated video objects and hence contain spatial and temporal information. Usually, hypervideo links have a visual representation on top of the video object to announce their existence to the viewer [7]. Until now research has primarily focussed purely on single video sources that include hypervideo links.

According to Cockburn et al.[8] revisiting information from the past is a common activity of users. In the web domain, standard browsers allow users to set bookmarks, use back and forward buttons or history lists in order to access previously visited web pages. Tools such as WebView[9] or Global Tree Browser [10] are examples that provide special functionality for revisiting Web pages. For local file access GoogleDesktop [11] and TimeScape [12] keep a history record allowing users to easily revisit their content. Based on the richness of context-sensitive video including multi-camera views the demand for a history tool enabling easy and fast access to previously selected video sequences has to be considered. Though, with time-based media, it requires more sophistication than for static webpages.

3 MiniDiver: Exploring Video Content on the Mobile Platform

The MiniDiver is designed to provide a user interface that is intuitive and responsive, yet powerful enough to take advantage of a context-aware video space on a mobile platform. The interface is based upon the concepts developed for the desktop MyView client, but with a focus on the iPhone™ which is a touch-enabled mobile device. There is also a personalized viewing mode where the user can browse video according to their unique preferences and interests. The MiniDiver requires options to display relevant metadata such as player positions, and names by overlaying them on top of the video content as in our ice hockey example. The MiniDiver supports the ability of users to control video with the usual transport control mechanisms (play, stop, backward and forward) but also provide mechanisms to change viewpoint, save and retrieve MiniDiving history and search/navigate using both episodic and semantic queries.

In this paper, we have prototyped our MiniDiver so that it uses local content residing on the iPhone™ or data over a network from a MyView server. The MyView server has a Director component that contains rules to feed video content to the MiniDiver appropriately, however, this is outside the scope of this paper. For communication with a simple networked client, the MiniDiver uses the Hive [13] communications protocol in order to stream content to the client device. Section 3.1 discusses the design of the MiniDiver that address the main four interaction mechanisms listed in Section 1.

3.1 MiniDiver: MyView Client Design

Our design concept of the mobile client is based on four user interfaces serving mobile context sensitive video. These interfaces are Content Browser, Video Player, Multi-Camera Browser and History Browser. In the following we will discuss and present each interface in more detail.

Content Browsing. Once the application loads the user will want to select some content to watch in the video player. We present the content selection interface using the usual iPhone™ table view hierarchy. In this commonly used interface style the user drills down into a series of lists that end in a detail view. In our navigation hierarchy the home screen allows the user to select a content source, and the next level allows them to select which events they would like to watch. A single event can be selected by tapping on it, and multiple events can be selected by tapping the navigation bars select button. This pops up an action toolbar at the bottom of the screen and transitions the list view into a selection mode. Events can then be selected by tapping on them and the action toolbar

Fig. 2. Modal bookmarks view activation animation

can be used to filter the result, or jump right into video playback. If the user has previously saved a bookmark they can access it from the startup screen by selecting the bookmarks button in the navigation bar. The bookmarks browser is presented as a modal view which slides up over the home screen. This was chosen to mimic the bookmarks functionality in MobileSafari (the iPhone™'s web browser). Tapping on a bookmark will take the user directly into the video player.

Once we enter the video player the system status bar is hidden to provide the user with a fully immersive experience. The interface is also re-oriented to landscape mode with the home button to the right of the screen. At this point we support only the landscape orientation for the Video Player and its siblings, the history browser and multi-camera browser.

Video Player. The video player screen, shown in Figure 1, contains much of the functionality of MiniDiver. From here users can navigate through time by grabbing the playhead or by using a two finger scrub gesture.

MyView content directly supports multiple video streams, so the user can also choose which camera angle they wish to view the action from. This is accomplished by a swipe gesture on the VideoPlayer screen. For example swiping left pushes the current video stream offscreen and brings in the stream of the camera physically located to right as shown in Figure 3. This gesture only allows for movement between streams in a one dimensional space, however we found that this was enough to spatially locate the streams in our test content. The one dimensional interface that we have developed is modelled after the photos application where users can swipe left and right to browse their photos. We have had to adjust this behavior slightly so that users can not slide a video stream

Fig. 3. Video stream switch animation. Interactive objects are highlighted in blue, and maintain consistency across views.

partially offscreen because we are limited by the mobile device processing power and the wireless network bandwidth to display two live video streams at once.

Notice that this approach only provides for relative camera view selection. That is, using the swipe interface you can move to cameras that are left or right (up or down if vertical swiping is added). Further, left and right camera views can be selected by the MyView Director service on the MyView server to provide meaningful interpretation as to viewing content more to the right or left. This can be done based on a quality-of-view analysis [14] or virtual viewpoints [15,16]. To have an absolute camera position selection requires meta-data tags that include a layout of the cameras in an absolute coordinate frame. We have not included this at this point since it is a variation on the Multi Camera Browser covered in Section 3.1, except we require mechanisms to show the video feeds from each of the absolute camera position. As well, like the relative camera position selection, the MyView Director will select a reasonably small set of either real or virtual cameras based on lowering cognitive load of the user, quality-of-view constraints and user preferences.

MyView content can be tagged with various information about objects that are onscreen and events that occurring at a given time. In our hockey example, content is tagged with player names and locations using hypervideo links and it also includes background masks. We use this information to allow users to select players in the video player view. This can be accomplished via a tap (single click), which triggers a player/object selection mechanism. Tapping the screen enters the selection mode which pauses the video and highlights all objects which are selectable. The user may then select an object by clicking on them; multiple objects may be selected in this manner. Video is started again by clicking on a non-selectable region of the screen.

The video player does have a simple translucent heads up display (HUD) and toolbar that can be activated by a single tap as shown in Figure 1. The HUD contains a play/pause button and a button to enter the multi-camera browser.

Fig. 4. Interface of Multi-Camera Browser

The toolbar contains a playhead and scrub bar, as well as a button to enter the history browser.

Multi-Camera Browser. Non-linear context sensitive video presentations based on a multi-camera scenarios require a visual disclosure of all available useful video streams to assist users to find their "best view". Hence, to allow users to easily navigate through available video streams we have created a multi-camera browser interface as shown in Figure 4.

We wanted the multi-camera browser to be a natural extension of the video player and we also anticipate that it will be used frequently. To best serve these requirements we made the multi-camera browser activate immediately when requested and only have a short activation animation. To achieve this we start loading visible video streams asynchronously right after kicking off the activation animation. Besides keeping initial load times down this also allows the browser to be fully responsive to touch input as soon as it is activated regardless of whether images are still being loaded from the network or disk.

Users access the multi-camera browser through a double tap on the HUD of the video player. When activated the application scales down the currently playing video to a grid view which also shows all of the other camera angles that are available. Video frames load asynchronously and fade in when ready, allowing the view to be interactive as soon as it is activated. The grid of camera angles can easily be scrolled with a vertical swipe gesture, and tapping on video frame will switch to that camera and transition back into the video player. If the user does not wish to select a camera there is a cancel button on the toolbar which will take them back to the video player.

History Browser. Context sensitive video content that includes multi-camera video stream requires new forms of navigation aid for its users. The History View allows users to navigate back to video sequences they had already selected in the past. To achieve this functionality, our mobile client keeps a record of every video scene users access combined with its start and end time of the video playback. With this data we are able to create an interactive History View Grid that can be used for navigation purposes. The History View Grid is shown in Figure 5.

The History View has two view modes: time-based and node-based. In both modes, the History Tree Structure is displayed with each clip being a rounded rectangle. Each rectangle is colored to indicate which camera the clip has video from, and the first frame of the clips video is also displayed to assist with differentiating cameras. The default state of the History View interface is the time mode.

In the time mode, the width of the rectangles are determined by the length of each clip while in the node mode the clips all have equal width. In both the heights are all the same, just big enough for a finger to tap. To switch to a previous clip or sequence, the user taps on the location in the history tree where they want to go to. The tree will animate to its new structure and keep the current sequence up to date in the player. So when a user taps the middle of a clip, that clip is made active and its sequence is loaded into the player. The playhead is also moved to the point in time corresponding to the location of the tap in the clip.

There are two axes in the default time view, a sequence axis and a time axis. The y-axis is the sequence axis and indicates which sequence falls at that level of the history tree. The time axis displays seconds and is inline with the start times and lengths of the displayed clips. The time axis can be switched to a "depth" axis by tapping the toggle button (tapping the toggle button again switches back to the time mode). In the node mode, all clips are displayed as having equal lengths, and the x-axis indicates their depth in the tree. This mode is useful if there are a large number of cuts over a short period of time, making them difficult to see on a scale proportional to time. As a history tree becomes larger, some portions will move off screen but can be reached just by dragging the tree until the desired sections are visible. The axes follow this motion and adjust their values accordingly, so the user always knows where they are. The

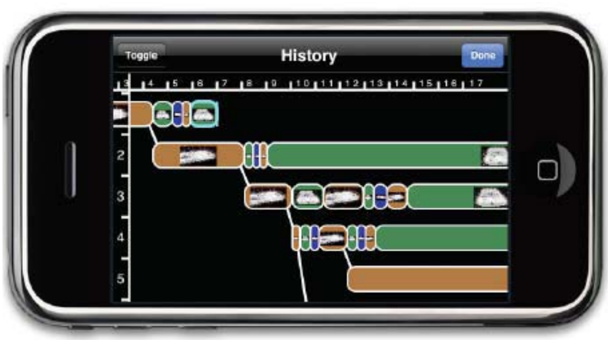


Fig. 5. History Browser Interface: horizontal axis represents the absolute time-scale of the synchronized videos; vertical axis represents branches showing user interaction

active clip is indicated by a blue border, and the playhead position by a blue vertical line on the active clip.

3.2 MiniDiver and MyView Communication Architecture

The MiniDiver uses the same underlying communication architecture as the MyView system, a flexible and modular framework called Hive[13,17]. All the components of MyView are based on Hive to facilitate simple communication and re-use of components on various platforms. Hive provides a communication protocol and an abstracted transport service for the transfer of data to different devices. The transport service can be implemented using various technologies (such as shared memory, ethernet or Bluetooth) to increase the number of devices that can communicate in the MyView framework. The connection paradigm is peer-to-peer to increase bandwidth but still allow centralized control.

The complicated processing of data (compression, background subtraction, player tracking) is performed centrally on a group of machines and the results are sent to the MiniDiver, using Hive as the transport medium. Each component is defined as a Hive *drone*, which is a module capable of performing a task and is controlled by an *application*. Applications have ultimate control over the drones in the system, and can construct pipeines by connecting drones together; the MiniDiver acts as a Hive application, controlling the drones and their connections. In future the MiniDiver will connect to a single point to receive data, to accommodate multiple MiniDiver clients.

3.3 Video Streaming

Getting video onto the MiniDiver is accomplished by streaming footage to the device using Hive for transportation. Multiple camera footage is stored centrally on a video server accessible via a drone operating as the interface between the video database and other Hive modules. Data transfers can operate in various modes, one of which is called the *request* mode. In this mode, data is only sent when requested by the destination drone; using this mode the images streamed to the device can change the camera footage currently being streamed by changing the request. This results in a lower latency than using a continuous streaming mode, and is scaleable to a larger number of camera sources.

Due to memory constraints on the mobile device, previous frames are not cached, so scrubbing along the time bar or navigating the history results in new requests for the previous data.

4 Results

This section presents the formats used for video streaming and the results of tests on latency and achievable frame rates on the mobile platforms of the iPhone and the iPod Touch.

The table in Figure 6 presents the quality of video used in the tests and the bandwidth required for each format. Due to the lack of support for developers

Content	Resolution	Format	Frame Rate	Data Rate
High Quality	480 × 270	JPEG Sequence	30Hz	630 KB/s
Low Quality	240 × 135	JPEG Sequence	30Hz	248 KB/s
Low Quality	240 × 135	JPEG Sequence	15Hz	124 KB/s
High Quality	480 × 270	H.264	30Hz	55 KB/s
Low Quality	480 × 270	H.264	15Hz	25 KB/s

Fig. 6. Comparison of data formats and rates - the iPhone$^{\mathrm{TM}}$ supports H.264 which is much more efficient, but not directly accessible to the developer. Our application employs JPEG sequences for compressed video streaming.

Content	Location	Frame Rate	
		iPhone$^{\mathrm{TM}}$ 3G	iPod Touch 2G
High Quality	Local	5 FPS	12 FPS
Low Quality	Local	17 FPS	30 FPS
High Quality	Network	×	8 FPS
Low Quality	Network	×	23 FPS

Fig. 7. Video playback performance using the JPEG sequence format (network testing performed only on the iPod Touch 2G)

to use the native video formats (H.264) the MiniDiver uses compressed JPEG image sequences for video. This also allows for frame-level requests from the video servers, which would be more complicated with H.264 because keyframes are widely spaced. H.264 formats are presented in the table to provide a comparison of data rates used.

Based on the tests we performed on the iPhone$^{\mathrm{TM}}$ 3G and the iPod Touch 2G there are obvious performance differences, shown in Figure 7. The iPod Touch 2G has an upgraded ARM processor from previous revisions, which explains its increased performance. Generally, using high resolution JPEG sequences results in low frame rates, with the iPod having the only acceptable rate for local access to video.

Network-based tests were performed only on the iPod using the request mode outlined in Section 3.2. There was a substantial drop in performance using the network, although we believe this could be significantly improved with optimization of the Hive implementation on the device. The MiniDiver performs well with the lower quality content either from local content or over the network. The latency in requesting a different camera angle is also reduced with lower bandwidth footage.

5 Conclusion and Future Work

We have created a new context-sensitive video browsing environment for the iPhone^{TM} called a MiniDiver. It provides mechanisms to select multiple camera views, navigate hyperlinked video, save and retrieve complex MiniDiving history, and some episodic and semantic search mechanisms. These mechanisms are necessary to allow people to experience the enormous amount of video content that is being generated. We have shown how multiple synchronous camera views can be accessed using a touch interface. We have included a mechanism to allow users to select moving objects in the video scene and get hyperlinked data associated with them. We have included a mechanism to support users traversing their history in complex, non-linear ways that re-interprets what it means to rewind and fast-forward in video. Our implementation of the episodic and semantic search mechanisms remains to be refined as we only provided semantic mechanisms and basic episodic mechanisms as found in the camera views in the MiniDiver. Our Hive implementation has provided performance measures to illustrate that delivering rich, interactive media data on a mobile device is feasible.

All of these mechanisms have yet to be thoroughly user tested as we have just defined the requirements needed in a video browsing environment to deal with the complex demands of large-scale, meta-tagged video. We continue to explore improved searching, enhancements for selecting one-to-many hyperlinked video data and developing means for the MyView server to supply only relevant data to the user based on their queries, context and profile.

This paper described our current prototype of a MiniDiver on the iPhone^{TM}. We are pursing design guidelines and prototypes for mobile devices for dealing with the anticipated changes in video content that are emerging due to the proliferation of video capture devices, video processing techniques and personalized video content. We believe a shift is needed for users to be able to effectively manage and share the wealth of video experiences the will have in the future. The MiniDiver provides some of the functionality that addresses this shift.

Acknowledgments. This research is funded by Bell Canada Inc. and the Natural Sciences and Engineering Research Council (NSERC), Canada. We thank all the members of the MyView team for their contributions.

References

1. Lou, J.G., Cai, H., Li, J.: A real-time interactive multi-view video system. In: MULTIMEDIA 2005: Proceedings of the 13th annual ACM international conference on Multimedia, pp. 161–170 (2005)
2. LeoAnnotation2008: Real-time multiview analysis of soccer matches for understanding interactions between ball and players. In: CIVR 2008: Proceedings of the 2008 international conference on Content-based image and video retrieval, pp. 525–534. ACM Press, New York (2008)
3. Grau, O., Hilton, A., Kilner, J., Miller, G., Sargeant, T., Starck, J.: A free-viewpoint video system for visualisation of sport scenes. SMPTE Motion Imaging, 213–219 (2007)

4. Lu, W.L., Okuma, K., Little, J.J.: Tracking and recognizing actions of multiple hockey players using the boosted particle filter. Image Vision Comput. 27(1-2), 189–205 (2009)
5. Goldman, D.B., Gonterman, C., Curless, B., Salesin, D., Seitz, S.M.: Video object annotation, navigation, and composition. In: UIST 2008: Proceedings of the 21st annual ACM symposium on User interface software and technology, pp. 3–12. ACM Press, New York (2008)
6. Shipman, F., Girgensohn, A., Wilcox, L.: Authoring, viewing, and generating hypervideo: An overview of hyper-hitchcock. ACM Trans. Multimedia Comput. Commun. Appl. 5(2), 1–19 (2008)
7. Stahl, E., Zahn, C., Finke, M.: How can we use hypervideo design projects to construct knowledge in university courses? In: CSCL 2005: Proceedings of th 2005 conference on Computer support for collaborative learning, International Society of the Learning Sciences, pp. 641–646 (2005)
8. Cockburn, A., Mckenzie, B.: What do web users do? an empirical analysis of web use. International Journal of Human-Computer Studies 54, 903–922 (2001)
9. Cockburn, A., Greenberg, S., McKenzie, B., Jasonsmith, M., Kaasten, S.: Webview, a graphical aid for revisiting web pages. In: OZCHI 1999: Proceedings of the 1999 Australian Conference on Human Computer Interaction (1999)
10. Killam, B.: A study of three browser history mechanisms for web navigation. In: IV 2001: Proceedings of the Fifth International Conference on Information Visualisation, Washington, DC, USA, p. 13. IEEE Computer Society, Los Alamitos (2001)
11. GoogleDesktop, http://desktop.google.com/
12. Rekimoto, J.: Time-machine computing: a time-centric approach for the information environment. In: UIST 1999: Proceedings of the 12th annual ACM symposium on User interface software and technology, pp. 45–54. ACM, New York (1999)
13. Afrah, A., Miller, G., Parks, D., Finke, M., Fels, S.: Hive: A distributed system for vision processing. In: Proc. of the Int. Conf. on Distributed Smart Cameras, September 2008, pp. 1–9 (2008)
14. Shen, C., Zhang, C., Fels, S.S.: A multi-camera surveillance system that estimates quality-of-view measurement. In: Proceedings of The IEEE International Conference on Image Processing (ICIP 2007), pp. III–193–III–196 (2007)
15. Miller, G., Starck, J., Hilton, A.: Projective surface refinement for free-viewpoint video. In: Proc. 3rd European Conference on Visual Media Production, IET, November 2006, pp. 153–162 (2006)
16. Zitnick, C., Kang, S., Uyttendaele, M., Winder, S., Szeliski, R.: High-quality video view interpolation using a layered representation. In: SIGGRAPH, pp. 600–608 (2004)
17. Miller, G., Afrah, A., Fels, S.: Rapid vision application development using hive. In: Proc. International Conference on Computer Vision Theory and Applications (February 2009)

Children's Choice of Games: The Influence of Prosocial Tendency and Education-Level

Vivian Hseuh-Hua Chen[1], Weirong Lin[1], Chiew Woon Ng[1], Su Li Chai[2], Angeline Cheok Eng Khoo[2], and Henry Been-Lirn Duh[3]

[1] Wee Kim Wee School of Communication and Information, 31 Nanyang Link, Nanyang Technological University, Singapore 637718
{chenhh,wrlin,chiewwoon}@ntu.edu.sg

[2] Psychological Studies Academic Group, National Institute of Education, 1 Nanyang Walk, Singapore 637616
{suli.chai,angeline.khoo}@nie.edu.sg

[3] Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117576
{eledbl}@nus.edu.sg

Abstract. This study employed the uses and gratifications approach to examine children's choice of gaming genres. The measure of prosocial behavioral tendency was used as an approximation of a child's offline gratification, and this was related to the exposure to three different genres of games (violent, aggressive and prosocial). The influence of education level was also taken into consideration. Data was compiled and analyzed from a survey conducted on Singaporean schoolchildren (N = 2,640). Overall results supported the supplementary model of gratification seeking behavior. Children with higher prosocial scores spent significantly less time playing violent and aggressive games, whereas children of a higher education level spent more time playing games of all genres. The results are presented and discussed.

Keywords: Uses and Gratifications, Games, Prosocial Orientations.

1 Introduction

With the fast increasing popularity of games, scholars are now beginning to take note of gaming as a media phenomenon deserving of their attention. Though more focus has been paid to the weightier issues such as game addiction and the negative effects of games, there is now a widening of the scholarly scope to include other issues such as motivations for gameplay and the social, economic and cultural aspects of virtual communities.

Besides improving the marketability of games, research into gameplay motivations and choice of games also enables a greater understanding of the appeal of games and how gameplay can be seen as an extension or a supplement to the real-life activities of gamers. Children are especially perceived to be easy prey for game developers, who market addictive virtual play-worlds that are targeted at specific age ranges. However, though it would be simple to portray children as largely passive consumers of games,

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 110–119, 2009.

the reality might not be so. There is a need to approach the consumption of games by children through a perspective that treats them as active consumers of media, which this study seeks to do via the uses and gratifications approach. To begin, the following literature review first traces the development of media theory, documenting the shift from the notion of a passive audience to that of an active one.

2 Literature Review

2.1 Media Effects Models and the Uses and Gratifications Approach

Early models of media effects emphasized the passivity of audiences, with the development of theories such as the hypodermic syringe or magic bullet theory. These perspectives viewed media consumption as an activity whereby the media exerted a unilateral influence on its viewers. Hence, its effects were conceived as something similar to how a syringe injects ideas directly into the consciousness of the masses.

Later on, developments in media effects theories shifted the prevailing consensus to give greater prerogative to individual agency. This newer approach could be traced back to the Lazarsfeld studies on media effects, which concluded that the media seldom had strong effects [1]. Following this, other scholars began rethinking the dominant, strong effects paradigm. Katz and Foulkes trace the development of the uses and gratifications approach from the 1940s onwards, noting how uses and gratifications research really became fruitful once researchers began to understand individual media use by turning to the social-psychological attributes of individuals [2]. These perspectives later matured into what is now known as the uses and gratifications paradigm, which represents a new way of thinking about the influence of media. Sharing somewhat similar premises with the psychological perspective of rational choice theory, the uses and gratifications approach emphasizes how individuals consume the media with the intention to obtain certain kinds of fulfillment.

However, scholars have also drawn attention to the potential drawbacks of the uses and gratification approach. For instance, some note that there is not yet a unified understanding of the approach, and many central concepts of the theory have not been made clear [3]. Others state that by focusing on audience consumption, the approach tends to be too individualistic [4]. Despite these limitations, media scholars suggest that these models still provide a useful framework from which to study communication, especially the issues surrounding the Internet and new media [3, 5, 6].

2.2 A Uses and Gratifications Approach to Gameplay

Employing a uses and gratifications approach essentially involves analyzing media choice by focusing on the notion of gratifications, or needs, that media use satisfies [3]. By identifying the motivations and traits behind individuals' choice of a particular media, the effects of media can then be viewed as a function of its use [7]. The perspective has been fruitfully employed in a range of media studies, especially those focusing on television usage. Yet, studies that have explicitly applied the approach to understanding gaming behavior have been few. Sherry and colleagues traced the studies on gaming that have used the uses and gratifications approach in the past [7]. According to the authors, Selnow's study [8] on gaming that adapted a television uses and gratifications scale was probably one of the earliest to investigate gaming practices. Their factor analysis yielded five key points, that (1) gameplay was preferable to human companionship, (2) it taught about people, (3) it provided companionship, (4) activity, and (5) escape. Another study by Wigand et al. [9] employed a similar approach found that gameplay was desired as it provided the necessary gratifications of excitement, satisfaction of doing well, and tension-reduction. More recently, studies by three groups of researchers [10- 12] have revealed gratifications such as passing time, enjoyment, skill testing, stress reduction, and competition [7].

Scholars have also used the uses and gratifications approach to link online and offline behavior. Scholars describe what is known as a "deficit model" (Steinkuehler, as cited in [13]) which refers to how the absence of a gratification in player's real lives drives them to seek it elsewhere. For instance, Burgoon [14] found that talk radio callers expressed less interest in communicating face-to-face, perceiving it to be less rewarding. It is possible that these individuals turned to talk radio to compensate for their lack of interest in face-to-face communication. In another study, individuals who perceived interpersonal interaction to be unrewarding turned to interaction based on Internet communication, hence perceiving it as a suitable alternative [15].

On the other hand, past research (e.g., of Lee and colleagues [13]) suggests instead that there may exist a *supplementary* model, where players are driven to seek more intense sources of gratification for a similar kind of need. The authors found significant positive correlations between gratifications obtained and sought in players' real and online (i.e., gaming) lives, thus supporting their claim. Similarly, Bruning [16] related motives for interpersonal and computer-mediated communication, finding that people enjoyed both interpersonal as well as computer-mediated-communication. This could be interpreted as an indication of how the fulfillment of a need in offline life drives individuals to seek other avenues of similar fulfillment, such as those available through online communication.

This strand of research that seeks to link offline and online gratifications seems promising, and the present study seeks to contribute to the debate by examining how real life gratifications are related to an individual's choice of game. This study uses the psychological measure of prosocial orientation as a measurement of the gratifications sought in real life. The prosocial orientation measure has been used in previous developmental psychology studies, and has been employed in several gaming studies as well. In addition to this, this study will also examine the effect of educational level on choice of game genre.

2.3 The Influence of Prosocial Orientation

Various psychological theories [17, 18] have proposed that basic human needs are universal, describing them as playing a central role in human life in which the fulfillment of these needs is critical in the process of growth, development and well-being.

Basic needs can be likened to a persistent feeling of hunger. For instance, Staub [18] conceptualizes basic needs as "the most fundamental motives" (p. 52) and proposes basic human needs as the driving force behind various courses of human action. In addition to the core set of basic needs documented in psychological literature, he emphasizes needs for "effectiveness and control" – the feeling of efficiency and control of events and situations; "positive identity" – positive self-conception and self-esteem (through enhancement of esteem from others); “positive connection to other individuals” – ownership of relationships which are positively connected to other individuals or groups; and “transcendence of the self” – the need to go beyond the self, which can all be satisfied by helping others (p. 56). For instance, if the act of engaging in helpful behavior has positive connotations in one’s culture and/or assists in the improvement of the welfare of others, not only would it be able to satisfy these needs of “positive identity,” “positive connection to others” and “effectiveness and control,” it would also be regularly affirmed within a community. This promotes an internalization of already developed values, beliefs and principles through social experiences, effectively leading individuals to cherish others’ welfare and view themselves as helpful persons. Hence, through the pursuit of basic human needs fulfillment, prosocial orientation and values are promoted, fostering the individuals’ prosocial behavioral tendencies.

It should be noted that the bulk of previous studies have viewed these prosocial behavioral tendencies as largely a consequence of, and not an antecedent to, exposure to gaming. For instance, an individual’s prosocial tendency is construed to be a result of his or her exposure to certain genres of games. While that interpretation is widely held, it is equally likely that these behavioral tendencies exert some form of effect on the choice of gaming genres. This study approaches the issue by taking prosocial behavior as chronologically antecedent to choice of game genre, and employs the psychological measure for individual prosocial behavioral tendency that is derived from Cheung et al’s work [19]. In particular, the helping behavior subscale will be used as an approximation of an individual’s gratification seeking behavior offline.

2.4 Education Level and the Development of Prosocial Orientation

In considering the effect of prosocial orientation on the choice of game genre, previous literature draws some attention to the influence of education level on the development of prosocial tendencies in children. Advancements in childrens’ ability to engage in moral reasoning are concomitant with the ability to see reality through the perspective of another, as well as the ability to grasp abstract concepts [20]. The development of prosocial behavior has also been documented to occur in tandem with advancement in education level and age of children. Eisenberg [20] described how elementary school children exhibit moral behavior primarily to seek approval or by adhering to stereotypical models of good. Throughout elementary school, direct reciprocity reasoning develops, where children exhibit good behavior as a means to self-gain (when others reply in kind). Finally, as late elementary school gives way to middle school and beyond, children begin to reason using morally abstract notions and they develop emotional responses to behavior (such as guilt), and acquire the ability to take on the perspective of others.

These differences in the moral development of the child affect the levels of prosocial behavioral tendency, especially in the key transition stage between late elementary school and early middle school. Development in moral reasoning may affect the degree of prosocial orientation, or at the very least alter the reasons underlying the prosocial orientation held to by the child. Hence, this study also takes into consideration the influence of level of education on the choice of game genre, through its effect on the development of the child’s prosocial orientation.

2.5 Research Questions

The present study examines the issue of gameplay using the uses and gratifications approach, by analyzing the links between individual prosocial behavioral tendency and the choice of game genres. This study will also examine the influence of the level of education on the choice of game. Hence, the following research questions are proposed:

RQ1: How does the degree of prosocial orientation affect the choice of game genre?
RQ2: How does the level of education of the child affect the choice of game genre?

3 Method

3.1 Participants

Participants were ($N = 2{,}640$) students from twelve primary and secondary schools in Singapore in which four of these schools were boys-only schools: 1329 primary school children (956 males and 372 females) and 1311 secondary school children (1004 males and 302 females). The age of primary school children ranged from 8 to 12 with the median age at 9 years. The ethnic composition of this sample was: 67.7% Chinese, 13.2% Malay, 8.6% Indian, 1.4% Eurasian, and 2.9% other. The median age for secondary school children was 13 and the ages ranged from 12 to 16. The ethnic composition for this sample was: 71.5% Chinese, 13.3% Malay, 6.9% Indian, 0.6% Eurasians, and 3.4% other.

3.2 Procedures

The questionnaires were administered to children collectively at school during March to May 2007. These questionnaires were counterbalanced in various orders. Children were measured on their prosocial behaviors and their video gaming habits.

3.3 Measures

An eleven-item subscale from *Prosocial Orientation Questionnaire* (POQ) [19] was used to measure children's *helping behavior*. Sample questions of *helping behavior* include "I would give up something to help my friends or family" and "I would volunteer to help a charity if they need my help". Children rated their agreement to these items on a scale from 1 (*strongly disagree*) to 4 (*strongly agree*). The scores were transformed to Percent of Maximum Possible (POMP) that ranged from 0 to 100. The primary school children mean sample for *helping behavior* was 70.86, standard deviation was 16.18 and the alpha was 0.73; while the secondary school children mean sample was 65.71, standard deviation was 13.30 and the alpha was 0.73. Children who scored above the mean were categorized as highly prosocial oriented, whilst children who scored below the mean were classified as low prosocial oriented.

Children were asked to list three of their favorite games and reported the estimated time they spent on each of these games (in hours per week). To measure the *prosocial content* of the game, two questions were included: "How often other players help each other in the game" and "How often they help others in the game," which were rated on a scale from 0 (*never*) to 3 (*almost always*). For the *violent content* of the game, another two questions were included: "How often do they shoot or kill creatures in the games" and "How often do they shoot or kill other players in the game" were assessed. To measure the *aggressive content* of the game, "How often do characters try to hurt each other's feelings in the game" was used. The prosocial, violent and aggressive content of game scores were computed through multiplying across each game genre's ratings and the amount of time spent playing the game. These scores were then averaged across all three aforementioned games [22] and were standardized.

This study chose to interpret game genre by seeking the respondents' evaluations of the activities they are involved within the game rather than allowing respondents to select from a list of pre-classified games. This gives the most accurate assessment of the specific activities that individuals are involved in as they play the game, as each game may provide a combination of several different types of gratification.

4 Results

The two research questions were tested through Multivariate Analysis of Variance (MANOVA), followed by a series of univariate F-tests to examine each of the dependent variables: Time Spent on prosocial (M = 1.35, SD = 2.29), violent (M = 0.89, SD = 2.02) and aggressive content games (M = 1.25, SD = 2.06). The results are reported in separate sections for each independent variable: Children's prosocial orientation, children's education level and interaction between prosocial orientation and education level.

Table 1. Descriptives of Prosocial Orientation and Education Level on Game Genres

	Prosocial Orientation				Education Level			
	Low		High		Primary		Secondary	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Violent Game	1.64	2.56	1.06	1.92	0.97	2.00	1.75	1.67
Aggressive Game	1.17	2.35	0.64	1.42	0.65	1.78	1.18	2.14
Prosocial Game	1.39	2.20	1.09	2.21	0.89	1.79	1.61	2.21

4.1 Prosocial Orientation on Choice of Game Genre

To test whether the degree of prosocial orientation affects the choice of game genre, a multivariate test was performed. The combined DVs (prosocial, violent and aggressive game content) were significantly affected by children's prosocial orientation, Wilk's Lambda, $F(3, 2125) = 13.35$, $p < 0.01$, $\eta^2 = 0.018$. High prosocially orientated children reported spending *less* time playing violent content games, $F(1, 2127) = 19.49$, $p < 0.01$, $\eta^2 = 0.009$. High prosocially oriented children also reported spending *less* time playing aggressive content games, $F(1, 2127) = 26.16$, $p < 0.01$, $\eta^2 = 0.012$. However, no significant result was found between prosocial orientation and prosocial content games.

4.2 Education Level on Choice of Game Genre

The MANOVA revealed that the combined DVs (prosocial, violent and aggressive game content) were significantly affected by education level, Wilk's Lambda, $F(3, 2125) = 20.80$, $p < 0.01$, $\eta^2 = 0.029$. Secondary school children reported spending more time playing all three genres of games as compared to primary school children [for *violent* content, $F(1, 2127) = 46.67$, $p < 0.01$, $\eta^2 = 0.021$; *aggressive* content, $F(1, 2127) = 24.61$, $p < 0.01$, $\eta^2 = 0.011$; *prosocial* content, $F(1, 2127) = 59.53$, $p < 0.01$, $\eta^2 = 0.027$].

4.3 Interaction between Prosocial Orientation and Education Level on Choice of Game Genre

The combined DVs were significantly affected by the interaction of children's education level and prosocial tendency, $F(3, 2125) = 2.95$, $p < 0.05$, $\eta^2 = 0.004$. There is a marginally significant interaction between children's education level and prosocial tendency on *aggressive* content game, $F(1, 2127) = 3.00$, $p = 0.08$, $\eta^2 = 0.001$. That is, though low prosocially oriented primary children ($M = 0.82$, $SD = 2.14$) spent more time playing aggressive content games than high prosocially oriented primary school children ($M = 0.53$, $SD = 1.48$), the discrepancy is even greater between low ($M = 1.40$, $SD = 2.46$) and high prosocially oriented ($M = 0.81$, $SD = 1.42$) secondary school children. No other significant interaction was found.

5 General Discussion

Results showed that children with higher prosocial scores tend to spend significantly less time playing violent and aggressive games. Relating this finding back to the models of gratification seeking behavior described in the review of literature, this finding seems to be supportive of the supplementary model. Children seek and receive a similar kind of gratification from their activities both in their offline lives as well as in the games that they play. Correspondingly, the children with lower prosocial scores tend to spend significantly more time playing violent and aggressive games, perhaps indicating that they also enjoy behaving less prosocially in-game, just as they do in their offline lives.

Results also revealed that secondary school children spent significantly more time playing violent, aggressive and prosocial games. Although in general primary school children reported themselves to be more helpful as compared to the secondary school children (refer to Measures under Methodology), their lower levels of prosocial gaming habits seem to suggest otherwise. In the present study, it was difficult to distinguish if these prosocial content games were single-player or multiplayer online games (such as Massively Multiplayer Online Role Playing Games). If we assume prosocial games to be largely multiplayer oriented, a predominant feature of such games is social interaction, where players are offered many opportunities to assist others and where the game dynamics require cooperation and coordination between players to achieve certain goals. Curiously, primary school children (who were generally more helpful) did not spend as much time in these prosocial games as secondary school children. A possible explanation lies in the deficiency in gratification that these younger children obtain by participating in such games. Previous studies have shown how maturity is a vital factor to successful community integration in this genre. Players often discriminate against younger players because they are perceived to be less mature, less willing to contribute to general goals, and less cooperative [23]. This fosters an unwelcoming communal climate for younger players and makes the gaming experience significantly less socially gratifying, which explains younger players' lower exposure to games in the prosocial genre.

In addition, some of the developmental theories on prosocial behavior suggest that young children are not fully developed in their comprehension of the concept of help. According to Cialdini's Socialization Model (as cited in [24]), children between the ages of 10 to 12 (i.e., primary school children) are not fully aware of the importance or emphasis of helpfulness in the society, and only help when told to do so. The fact that primary school children in this study on average reported themselves to be quite helpful (more so than those in secondary school) might be due largely to how the importance of certain helping behaviors are frequently impressed upon them by their parents and teachers. Since the measurement of prosocial behavior is measured in simple and clear-cut scenarios, with no moral ambiguity, these children find it easy to pick out the behavior that is required of them. For instance, the prosocial questionnaire had items such as "I always think of helping people who are disabled." However, in-game helping behavior often involves complex decisions in scenarios riddled with moral grey areas. Hence, it should come as no surprise that younger children are unable to fully grasp the nature of helping behavior that unfolds in online interaction. Again, this hinders their expression of prosocial behavior and makes their gaming experience less socially gratifying, which accounts for lower rates of prosocial game exposure.

Finally, it should be noted that though this study's use of the prosocial behavioral tendency provides a novel way of approximating gratification-seeking behavior, there is a potential danger of reading too much into the behavioral tendencies of respondents. In future studies, such psychological measures should be used in combination with more traditional and straightforward methods of measuring gratification seeking behavior to obtain a comprehensive measure of the concept.

Acknowledgments. The study was funded by a grant from the Singapore Ministry of Education and Media Development Authority (EP1/06/AK).

References

1. Rogers, E.M.: A History of Communication Study. The Free Press, New York (1994)
2. Katz, E., Foulkes, D.: On the Use of Mass Media as Escape: Clarification of a Concept. Public Opinion Quarterly 26, 377–388 (1962)

3. Ruggerio, E.T.: Uses and Gratifications Theory in the 21st Century. Mass Communication and Society 3, 3–37 (2000)
4. Elliott, P.: Uses and Gratifications Research: A Critique and a Sociological Alternative. In: Blumler, J.G., Katz, E. (eds.) The Uses of Mass Communications: Current Perspectives on Gratifications Research, pp. 249–268. Sage, Beverly Hills (1974)
5. Cho, J., Zuniga, H.G.D., Rojas, H., Shah, D.V.: Beyond Access: The Digital Divide and Internet Uses and Gratifications. IT & Society 1, 46–72 (2003)
6. Morris, M., Ogan, C.: The Internet as Mass Medium. Journal of Communication 46, 39–50 (1996)
7. Sherry, J.L., Lucas, K., Greenberg, B.S., Lachlan, K.: Video Game Uses and Gratifications as Predictors of Use and Game Preference. In: Vorderer, P., Bryant, J. (eds.) Playing Video Games, pp. 213–224. Routledge, New York (2006)
8. Selnow, G.W.: Playing Videogames: The Electronic Friend. Journal of Communication 34, 148–156 (1984)
9. Wigand, R.T., Borstelmann, S.E., Boster, F.J.: Electronic Leisure: Video Game Usage and the Communication Climate of Video Arcades. Communication Yearbook 9, 275–293 (1985)
10. Phillips, C.A., Rolls, S., Rouse, A., Griffiths, M.D.: Home Video Game Playing in Schoolchildren: A Study of Incidence and Patterns of Play. Journal of Adolescence 18, 687–691 (1995)
11. Griffiths, M.D.: The Observational Analysis of Adolescent Gaming in UK Amusement Arcades. Journal of Community and Applied Social Psychology 1, 309–320 (1991)
12. Vorderer, P., Hartmann, T., Klimmt, C.: Explaining the Enjoyment of Playing Video Games: The Role of Competition. In: Proceedings of the Second International Conference on Entertainment Computing, pp. 1–9. Carnegie Mellon University, Pittsburgh (2003)
13. Lee, H.C., Wong, S.Y., Wong, Y.Z.: The Massively Multiplayer Online Game in Singapore: Gamers and their Motivations. Final Year Thesis, Nanyang Technological University, Singapore (2006)
14. Burgoon, J.K.: The Unwillingness to Communicate Scale: Development and Validation. Communication Monographs 43, 60–69 (1976)
15. Papacharissi, Z., Rubin, A.M.: Predictors of Internet Use. Journal of Broadcasting and Electronic Media 4, 175–196 (2000)
16. Bruning, S.D.: An Examination of the Social, Psychological, and Communication Variables that Influence User Perceptions of Computer-Mediated Communication Technologies (Doctoral Dissertation, Kent State University) Dissertation Abstracts International (1992)
17. Maslow, A.H.: Towards a Psychology of Well Being. J. Wiley and Sons, New York (1999)
18. Staub, E.: The Psychology of Good and Evil: Why Children, Adults and Group Help and Harm Others. Cambridge University Press, New York (2003)
19. Cheung, P.C., Ma, H.K., Shek, D.T.L.: Conception of success: Their correlates with prosocial orientation and behavior in Chinese adolescents. Journal of Adolescence 21, 31–42 (1998)
20. Eisenberg, N., Miller, P.A., Shell, R., McNalley, S., Shea, C.: Prosocial Development in Adolescence: A Longitudinal Study. Development Psychology 27, 849–857 (1991)
21. Kohlberg, L.: The Philosophy of Moral Development: Moral Stages and the Idea of Justice. Harper & Row, San Francisco (1981)

22. Gentile, D.A., Anderson, C.A., Yukawa, S., Ihori, N., Saleem, M., Lim, K.M., Shibuya, A., Liau, A.K., Khoo, A., Bushman, B.J., Huesmann, L.R., Sakamoto, A.: The effects of Prosocial Video Games on Prosocial Behaviors: International Evidence from Correlational, Longitudinal, and Experimental Studies. Personality and Social Psychology Bulletin (in press)
23. Williams, D., Ducheneaut, N., Xiong, L., Yee, N., Nickell, E.: From tree house to barracks: The social life of guilds in World of Warcraft. Games and Culture 1, 338–361 (2006)
24. Dovidio, J.G., Piliavin, J.A., Schroeder, D.A., Penner, L.A.: The Social Psychology of Prosocial Behavior. Lawrence Erlbaum Associates, New Jersey (2006)

Player Experience Evaluation: An Approach Based on the Personal Construct Theory

Francesco Bellotti, Riccardo Berta, Alessandro De Gloria, and Ludovica Primavera

DIBE – Department of Electronics and Biophysical Engineering
University of Genoa, via Opera Pia 11 a,
16145 Genova, Italy
{franz,berta,adg,primavera}@elios.unige.it

Abstract. The scientific and industrial community related to the videogame (VG) research and business is ever more concerned about the need for proper evaluation and assessment of games. This paper proposes an assessment methodology based on the Personal Construct Theory (PCT). The PCT allows identifying constructs can be processed to define a space where domain-relevant items – VGs, in our case - can be positioned. The main praise of PCT is that the test-leading researcher does not supply users with a predefined set of constructs, which may bias the evaluation process. Moreover, PCT joins qualitative aspects with a quantitative evaluation of their relevance, which is particularly useful for an operational approach also to game design. In this paper, we study the application of the PCT to the particular case of the evaluation of whole typologies of VGs. Discussing the results, we draw and highlight that VGs are perceived as engaging challenges where personal abilities are continuously put to the test. This stresses the reactive nature of VGs and the fact that players like being stimulated and developing and testing their reaction capabilities.

Keywords: Videogames, User Experience, videogame testing and evaluation, Repertory Grid Technique, Personal Construct Theory.

1 Introduction

Videogaming has acquired a great relevance since years, not only for the youngsters. Serious games applications expand the horizons of games to include scientific simulation and visualization, industrial and military training, medical and health training and education, as well as public awareness and policy change [1].

In the Transformation Economy people is expected to pay to have experiences that are transformational [2], and games and simulations allow significantly extending the range of experiences a person can live, with an ever growing level of realism. Producing a videogame (VG) has become a huge and expensive project, since the market is characterized by trends toward hit products and fewer Stock-Keeping Units [3]. In this scenario, awareness has risen of the importance of evaluation methodologies able to support producers and evaluators in assessing products since early phases of a project. However, there are not commonly shared and accepted techniques, yet [4].

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 120–131, 2009.

In this paper we propose and discuss a VG evaluation methodology based on the Personal Construct Theory (PCT) of personality [5], which aims at determining an idiosyncratic measure of personality. The theory has been initially applied in psychological clinic interviews [6] and then in a number of other fields, also including website evaluation [7]. The PCT assumes that individuals perceive and evaluate the world through similarity-dissimilarity poles, called “personal constructs”. Such poles can be used to define an individual’s personal construct system relevant to the analyzed domain - in our case the evaluation of VGs. The main praise of PCT is that the test-leading researcher does not supply users with a predefined set of constructs, which may bias the evaluation process [8]. Rather, constructs are defined a posteriori, based on an analysis of the free user comments. This is particularly promising for VG evaluation, because it stresses the centrality and relevance of the player experience, which is a fundamental value for a game success. Moreover, the theory gives indications based on the subjects’ free evaluation of their domain (i.e. VG) experiences, which may give also ideas for new products, not only the evaluation of existing ones. PCT joins qualitative aspects - the elicited personal constructs - with a quantitative evaluation of their relevance, which is particularly useful for an operational approach also to game design.

We study the application of the PCT to the particular case of the evaluation of whole typologies of VGs (e.g. Sports, Racers, First Person Shooters). In particular, we intend to explore what the personal constructs are that define the space of the games - according to test player experiences - and observe how current game typologies are placed with respect to such dimensions (i.e. the constructs). Finally, we also try to see if new, “hybrid” game typologies could be devised based on the combination of important constructs.

The paper is organized as it follows. Section 2 presents the main general aspects of PCT, section 3 provides a literature review of VG evaluation methodologies. Section 4 describes the procedure we have defined for the evaluation of player experience. The section 5 discusses the results, proposing observations and findings.

2 The Personal Construct Theory (PCT)

Personal Construct Psychology (PCP), also referred as Personal Construct Theory (PCT), originated with the pioneering work of George Kelly [9]. Kelly’s theory of personality was based on the metaphor of the Inquiring Man [5]. A person is seen as a scientist who tries to make sense of the world by continually forming hypotheses about the world and experimenting and observing. The PCT assumes that the individual perceives and evaluates the world (e.g. other individuals, situations, etc.) through similarity-dissimilarity poles, called Personal Constructs. The meaning of any construct is defined by the complementarities of opposite poles, in Kelly words: “the differences expressed by a construct are just as relevant as the likenesses.” [9]. The personal construct is the elementary unit that a person uses during her process of construction of meaning about the world. Thus, constructs represent the personal view of the world the subject has constructed based on her own experience.

2.1 The Repertory Grid Technique

The Repertory Grid Technique (RGT) [10] is a methodological extension of PCT to empirically eliciting and evaluating individual's subjective personal construct system relevant to a topic. An RGT is a table in which rows contain constructs and columns contain elements (instances of the topic under investigation). To fill the table, the person refers to a rating system to quantitatively assign each element to one of the poles of each qualitative constructs. A grid is constructed for each subject participating in a test session. A subject produces his constructs list composed of bipolar dimensions that the person sees as important for talking about the elements. The elicitation process is based on the use of triads of elements. Each subject is presented with triads of such elements and asked to indicate in what respect two of the three are similar to each other and differ from the third one. These similarity/differentiation factors are recorded as bipolar constructs. Different triads may give rise to the similar (or even same) bipolar properties. Then, after having identified his individual constructs, the subject is asked to rate the degree to which each element is in relation to each bipolar construct, according to a Likert-type scale. The personal constructs can be then analyses and summed together, and the most frequent ones are assigned a higher weight as they represent the most appropriate properties able to characterize the domain. All such bipolar constructs (each one with a different weight) are thus elicited as the proper dimensions (or classification modes) along which the domain elements are evaluated by the user.

The main feature of RGT is that the researcher does not supply users with a predefined set of constructs, which may bias the evaluation process [10]. Rather, constructs are defined by the participants, based on the analysis (triad methods) of his comments.

3 Literature Review

The main feature of a successful VG is the enjoinment of the player, but there is no commonly accepted model for player enjoyment [4]. In literature there are several heuristics for designing and evaluating games, and all of them are focused on three main aspects: gameplay, mechanics and interface [11]. These heuristics are often fragmented and contradictory, because doesn't exist evidence about the relevance of used criteria versus VGs enjoinment [12]. From another point of view, many researchers have proposed models based on psychological theories, including disposition theory [13], attitude [14], transportation theory [15], para-social interaction [14] and flow [16]. The main drawback of these attempts is their predetermined structure [17]. Examples include methodology based on exploratory interview, which has the problem of imposition of terminology that prevents the participant to make use of their proper terminology in the discussions [18]. It is needed to combine these elements into a set of dimensions not predetermined a priori, but extracted from real users. Our approach, based on the RGT is focused on eliciting dimensions that are meaningful to the participant and not only to the experimenter. The data that we capture from end users has not been influenced by the researcher's theory. Moreover, it takes also into account users' idiosyncratic views in contrast to the other approaches, which impose a priori defined dimensions.

Although the RGT technique was initially developed in psychological clinical settings in order to determine an idiosyncratic measure of personality, it has been applied to a wide variety of cases, including evaluation of user experience [8], also in the user experience evaluation research [19] [20]. Based on this analysis and our experience, we believe that RGT facilitates an objective approach to capturing subjective aspects of VG evaluation by allowing an open, broad and sensitive analysis of the several aspects that might have an impact on the user's experience of a VG.

Others techniques attempt to model user enjoinment based on intrusive measurements of user's physiology [8]. Example of these are based on facial electromyography as a measure of positive and negative emotional states [21], or on the children's heart rate signals for capturing and modeling individual entertainment preferences [22].

4 Applying the Method to the Player Experience Evaluation

This section describes the methodology, based on RGT, which we have defined and followed for VG evaluation, while the next session presents the results. A test session consists of three major steps: initial focus (users are asked to think about their concrete experience with items in the investigated domain), construct elicitation (users have to identify similarities and differences among elements/items in order for researchers to identify bipolar constructs), rating (users have to rate each element along each elicited construct). We describe in detail the following steps, but first we describe the test user selection procedure.

4.1 Subject Selection

In order to test our RGT to VGs, we have recruited test subjects by contacting teachers of three different high-schools with scientific and technical background. To each school we proposed a one-day workshop about VGs. We tested a total of 64 students in an age range from 16 to 19 (mean 17,4; stddev 0,64). There were 13 females and 51 males. We asked for students with a strong interest for VGs, as we wanted to have opinions from experienced users, who have a good familiarity with state of the art VGs. This subject selection criterion should allow us to have a realistic analysis of the current VG panorama. Of course, a similar analysis could be conducted on a more generalist population sample, for example if we were interested in understanding on how to attract new typologies of users (e.g. women). But this is not the specific purpose of this work. In a pre-test questionnaire phase we had tested the users' level of familiarity with VGs, in order to check the appropriateness of our target group. As expected, participants were quite familiar with VGs (mean 2,56; stddev 1,15) and would usually spend a huge amount of time playing with consoles and PCs (1-3 hours per day).

4.2 Initial Focus

This preliminary phase invites test users to focus on the investigation domain, in particular by thinking of the elements. Elements are a selection of objects relevant to the topic/domain. In our analysis, we have chosen as elements 8 major game genres: Adventures, Racers, First-Person Shooters (FPS), Real-Time Strategy games (RTS), Sports, Arcades, Virtual Worlds (VW), and Role-Playing Game (RPG). The PCT highlights that a person makes sense of the world by experimenting and observing. It is very important that each participant focus on the genres using her personal experience. For this reason, participants are asked to select a well known VG for each element (genre) and to focus on it in the subsequent step.

4.3 Construct Elicitation and Ratings

We apply the triad method. Every participant ias interviewed by a researcher. Users are asked to identify one VG for each genre defined in the previous step. During the interview, the participant is invited to think about the 8 VG genre samples in groups of three – this is the triading, in RGT's jargon. For every meaningful triad, the participant has to identify a property or a quality important enough to explain how two of the elements are similar, but different from the third one. The user has to put a label on that property. For instance, in the group of Quake (FPS), Warcraft (RTS), and Colin McRae Rally (Racers), one of the participants picked out Colin McRae, and described the property as "the Colin McRae situations, 3D world, settings and simulation are very realistic and credible, while Quake and Warcraft are fantasy", that we abstracted and added to the "likelihood" construct. This abstraction and cataloguing phase – namely, elicitation of the constructs - is described in detail in the 5.1 subsection.

For each one of the elicited constructs, the participant rates each one of the 8 elements (also those that not appear in the specific triad from which a particular construct was elicited) according to a five-graded scale, where 0 represents a low relevance of the property for the games in the triad, while 4 represents a high relevance.

5 Evaluation of Videogame Player Experience

This section presents the results obtained by the above presented RGT method. The data collected in the experiment consist of the personal constructs elicited by the participants and the rating of the investigated elements (VG genres) along such elicited constructs. The first data provide a qualitative insight about the characterization – according to the user experience – of the investigated domain. The second data provide a quantitative measure of the degree of relevance of each elicited construct. This section describes the methodology for processing such data.

5.1 Elicited Dimensions

\RGT typically produces a large number of individual tables (we have collected a total of 570 constructs). Sample bipolar constructs include: "a good experience is necessary to play – it can be played without training", "the game requires cognitive effort – it can be played also by a monkey". Differently than in clinical studies, we are not interested in the identification of idiosyncrasies. Instead, our research aims at finding out the basic and commonly shared criteria that players use in evaluating VGs. In our study each table contains a fixed number (8) of elements, a fixed number of constructs (we ask participants to consider ten triads) and a shared rating system (a bipolar five-grade scale). From these shared features, it is possible to apply statistical methods (semantically blind methods) and/or semantic similarity (statistically blind methods) to analyze data interpersonally. In particular, we can search for patterns to compare constructs from all participants and assemble constructs showing some degree of similitude into groups.

We have analyzed the collected data and grouped the users' personal constructs into clusters. To this end, each pole of each construct was carefully semantically interpreted (using a discourse analysis technique) to capture the character of construct and to propose some shared labels – called dimensions - able to characterize all similar construct as a whole. The number of test users' personal constructs that have been grouped under a dimension is an indicator of the relevance of that dimension among the subjects (we call this the dimension weight W). The higher is the number of grouped constructs, the

Table 1. Description of dimensions

Dimensions	Descriptions	W
Ability demand	The game requires user's ability in terms of capabilities, skills and cognitive efforts.	94
Dynamism	The activities in the game space are varied, unexpected and with frequent and quick changes.	58
Style	The game has a distinctive appearance (modern, old) on the basis also of its genre (a F1 race game should recreate a Grand Prix' atmosphere)	48
Engagement	The game is exciting, challenging, charming, compelling. The player is drawn never to stop the game.	38
Emotional affect	The game affects the player's emotional state. It induces anxiety, excitement, and suspense.	35
Likelihood	The game situations, 3D reconstructions are very realistic and credible.	33
Sociality	The game mechanic allows the player to form social groups.	28
Enjoyability	The game provides amusement, satisfaction, pleasure and gratification.	28
Complexity	The gameplay is composed by a huge amount of entangled experiences; the virtual word to be explored is wide and with a great variety of landscapes.	24
Playability span	The game has the ability to involve players for long periods of time (e.g. several months).	23
Technological Quality	The game is very accurate from the technological point of view: good graphics, sound, multimedia contents, GUI design.	23
Plot	The game is fascinating because the plot is very intriguing.	21
Physical effort	The game is very demanding in terms of physical activities (stress, muscular tension, movements of muscles).	20
Replayability	The game experience is varied enough (in terms of plot or mechanic) for the player to repeat it several times without exhausting all possibilities.	18
Controllability	Players are given large freedom in selecting actions, interacting with the ambient and configuring options. The player's avatar is highly customizable, also concerning its personality.	15
Identification	The player controls a character with a specific role in the game plot and she/he feels a sense of identification with the character.	13
Open-end	The game plot is almost endless. The player is involved in long stories that continue indefinitely.	12
Active involvement	The player is an active subject in the game and she/he plays also a creative role. For example, she can build her own 3D models in the virtual world, or modify the character's aspect with personal contents.	10
Competition/ cooperation	The game gives high relevance to the score. It offers possibilities to cooperate or compete with other players towards a common goal.	10

more relevant is the cluster and higher is its weight, as more test users have highlight it. To reduce the risk of possible biases or misunderstandings in the definition of dimensions we have selected labels among existing personal constructs rather than inventing new ones. As an example of this labeling work, the dimension “Ability” takes into account 94 personal constructs. We have selected the term “ability” because it is the most frequent used in the participants descriptions of the related constructs. The 23 dimensions that resulted can be regarded as the most pertinent features of the participants' understandings of the VGs genres considered in the study and a list with the most important dimensions is provided in the table 1.

Analyzing the quantitative data on dimensions, we can gain significant insights into factors that players consider key in characterizing their gaming experience.

First, we see that VGs are perceived as engaging challenges where personal abilities are continuously put to the test, in particular with emotional involvement and frequent and quick changes in context and situation. This highlights the reactive nature of VGs and stresses the fact that players like being stimulated and testing/showing/training their reaction capabilities, also in a mechanical – almost non conscious – way, as it emerged from some informal comments we recorded during the (e.g. “I play to distract myself”, “When I play I don't want to think”). It has to be noted that, despite users declare they like variety and change, they do not explicitly report the repetitive nature of several of the most popular games (e.g. shoot'em up and some sport games), which is apparent to the researchers. We may argue that players like intrinsic repetitive patterns (e.g. shooting, firing, jumping, running, etc.) but they want them to be applied to several different contexts and situations, and at different levels of difficulty. We may assimilate this to a basic, easy to learn alphabet that has to be used in a variety of contexts to tell a variety of tales.

Plot and sociality are not so important. This highlights that users are not particularly interested in these aspects and that the currently most successful VGs do not strongly support these dimensions.

Distinctive style and multimedia quality are relevant evaluation parameters, which highlights the importance of design and technological excellence.

The dimensions that we have elicited from the user tests tend to include the heuristics normally used by professionals and players when writing game reviews in specialized magazines and web sites (e.g. [23]), that are quite focused on the game interface and mechanics [24]. Other heuristics are pointed out in academic works, such as the well established GameFlow approach [25] [26], and underline the importance for a game to involve the player through immersion and concentration-keeping challenges. The dimension list from our tests tends to include the heuristics of both the above mentioned typologies and has two important features: it gives a weight to the relevance of the various dimensions and is based on the experience of real users rather than on psychological theories or on the practice and knowledge of a professional user or reviewer.

5.2 Assessing Videogame Genres along the Elicited Dimensions

The subsequent step of test consists in the users rating in a 0-4 scale of each VG genre along each one of the elicited dimensions. The data collected from each single test participant is then summarized in a table – namely, the Repertory Grid - that reports the average value of the personal tables. Table 2 shows the RG for our case.

Table 2. Repertory Grid for the evaluation of VG genres along the elicited dimensions

	Adv	Racer	FPS	RTS	Sport	Arcade	VW	RPG
Sociality	2,50	2,89	2,93	3,38	4,00	2,26	2,83	2,56
Open-end	1,00	3,56	2,75	3,00	3,20	2,67	1,75	3,50
Likelihood	2,31	3,06	3,29	2,89	3,14	3,00	4,00	2,72
Dynamism	3,79	3,51	3,64	2,64	3,36	3,62	3,48	3,12
Ability Demand	3,14	2,88	3,43	3,54	3,20	2,80	2,74	3,51
Emotional Affect	3,06	3,04	3,55	2,71	2,46	2,81	2,33	3,79
Active Involvement	3,14	2,20	4,00	3,00	2,57	2,43	3,67	4,50
Competition Cooperation	1,00	2,14	3,44	3,13	3,75	2,38	3,00	3,00
Identification	3,44	3,00	2,50	1,80	3,08	3,43	4,75	3,33
Plot	3,31	2,62	3,27	3,67	2,67	1,82	3,00	3,43
Re-playability	2,33	3,13	3,38	3,71	3,33	2,88	3,67	4,08
Playability Span	3,60	3,24	3,33	3,45	3,86	2,93	4,00	3,45
Controllability	2,50	3,00	2,88	3,33	3,43	2,33	3,67	3,33
Complexity	2,45	3,20	2,94	3,17	3,25	2,55	3,60	4,15
Style	2,75	3,59	2,77	3,04	3,16	2,45	2,88	3,45
Tecnological Quality	2,62	3,44	3,82	3,29	3,62	2,33	4,55	3,30
Physical Effort	3,17	3,43	3,10	3,83	3,12	2,80	3,22	3,14
Engagement	3,67	3,48	3,53	3,81	3,08	3,16	3,43	3,71
Enjoyability	3,67	3,91	3,21	3,50	3,43	4,14	3,44	3,60

5.3 Results at a Glance

While, the RG table provides the detail of the quantitative analysis, an alternative presentation method may be used, especially in order to provide a more immediate visualization of the results. This method exploits a radar graph. The graph is consists of a two-dimensional chart with several equi-angular axes - called radii - one for each dimension. An element (VG genre) is represented by its values shown on each radium of the graph. For each element, a line is drawn connecting the values, obtaining a star-like graph. If the number of elements is large, the graph tends to become cluttered, especially if printed in black and white. Figure 1 reports the radar graph for just 4 elements of our analysis (FPS, RPG Adventures and VW). However, the color graph representing all the 8 investigated elements is interesting in order to have a complete outlook of the domain, and is available online[1].

As anticipated, the graphical representation is particularly useful to grasp an immediate idea of the analysis, suggesting immediate answers on issues that are relevant to game designers and evaluators, such as: what dimensions are prevalent for a given genre? What genres rely on a particular user construct? Which genres are similar/different among each other, and why? Are there some outliers? We will discuss these questions in the next subsection.

5.4 Interpretation of the RG Results

Observing the radar plot it is possible to get information about the single elements or identify clusters of game genres with similar features. For instance, looking at the star plot of the Virtual World genre we can see that is characterized by a star plot with

[1] www.elios.dibe.unige.it\radar.jpg

peaks in single dimensions. It is the most technological demanding genre, is rated below the average concerning the physical effort and scores high in playability span, identification and likelihood. Other genres, instead, have a more balanced star graph. We expect that genres with a balanced graph tend to provide the player with a more "complete" experience. In effect, FPSs, that are the most common and popular games at present – in the sense that they are played by almost every videogamer, even if they are not necessarily the most appreciated type - have a graph without particular peaks. On the other hand, genres with a sparked star plot (e.g. VWs, RPGs) are likely to be particularly appealing to more specialized types of player.

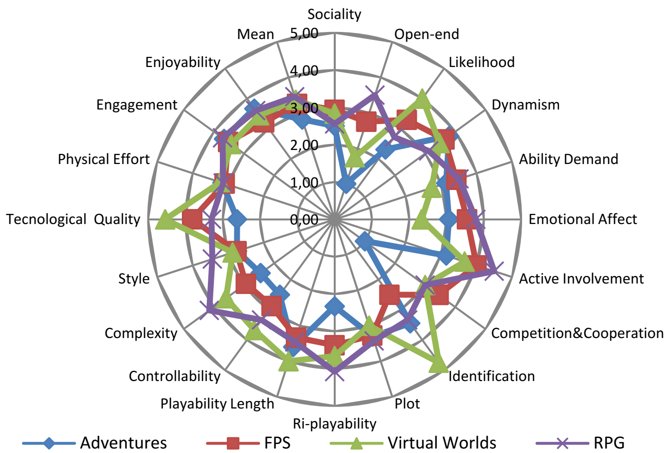


Fig. 1. The radar plot of games genres against genres

In an adventure VG, the player apparently likes plunging her/himself in the adventure defined by the author. In fact, the plot dimension has a significant value. Correspondingly, the adventure genre has the lowest score in the "open-end" dimension. We can argue that, for an adventure, an intricate and meaningful story is the most important feature and it is difficult for a game to combine such a type of plot with an open-end. Likelihood, on the other hand, is not a decisive factor for an adventure - in fact we have a number of successful adventures that involve fantasy settings. While likelihood is important for VWs, probably because the users like having clear, familiar references in a VW where a pre-defined, authored plot is not present and players have to build meaning and get enjoyment from their interaction among each other and with the environment. We argue that results may suggest that the lack of plot has to be compensated also by high technological quality (multimedia and GUI) as well.

Observing the single dimensions, we can observe, that the most important one – according to the dimension analysis in tab. 1 – namely "ability demand", is characterized by a very little variance among the elements. This means that all types of VGs have to be strong in such dimension in order to appeal to the audience. This is true also for all the other important dimensions, such as dynamism, style, engagement. This spurs us to argue a generalization. That is that all the VGs, independent of their genre, have to provide some main ingredients to be successful. VG design has few niches, suggested by peaks in the graph.

Player perception of RPGs stresses the importance of complexity and active involvement. In fact, state of the art RPGs are characterized by a huge number of rules, cards and objects. The player has a lot of things to understand, learn and remember.

For VWs, the identification dimension is quite significant. This is reasonable since VW players have to differentiate among each other to be distinct in the VW. This is also already supported by some state of the art features, such as the customization of the avatar through the choice of glasses, hat, cloths, character aspects, etc. We have already seen some important distinctive factors for VWs (likelihood, technological quality, long span playability, identification). Generalizing, we can see that VWs (and RPGs, to a lesser extent) tend to require some more specialized features for game design.

Combining information from the dimension analysis (table 1) and the radar graph is interesting to think of new "hybrid" game genres. In particular, we try to combine game genres that have peaks in the most important dimensions – given for granted that, as we have seen, the most popular dimensions (e.g. "ability demand") are genre-independent and thus have to be targeted in any type of game design. Dynamism (the second highest weight in table 1) is important in particular for adventures, while style (the third weight) for racers and RPG. Combining adventures and RPGs could be of interest even if it is very challenging from an authorial point of view, since the game should combine a strong, compelling and meaningful plot (typical of adventures) with the possibility for every player of freely and actively interpreting her/his own role.

6 Conclusions

The scientific and industrial community related to the VG research and business is ever more concerned about the need for proper evaluation and assessment of games [4]. Devising suited methodologies is difficult, because the gaming experience is quite subjective and involves a number of different and possibly also conflicting aspects. This is apparent, for instance, as different game genres are now popular, that target different types of audience.

In this paper we have proposed an assessment methodology based on the PCT psychological theory and evaluated its validity in a test aimed at defining the features of a space that defines how state of the art commercial games are perceived by players. Results highlight that VGs are perceived as engaging challenges where personal abilities are continuously put to the test. This stresses the reactive nature of VGs and the fact that players like being stimulated and developing and testing their reaction capabilities. The space defined from the our tests' dimension tends to include the heuristics of the two major typologies of game review/evaluation methodologies: the approach typically used by professionals and players, who are quite focused on the game interface and mechanics, and the academic work that stresses the importance of an immersive experience and of concentration-keeping challenges. PCT results add two important features. First, it provides a quantitative evaluation, by giving a weight to the relevance of each dimension. Second, it is based on the experience of real users rather than on psychological theories or on the practice and knowledge of experts.

We have then analyzed how current game typologies/genres are positioned with respect to such dimensions. This led to a number of interesting characterizations. Generalizing, we observed that some typologies cover a number of dimensions (e.g. FPS), while other typologies have peaks in single dimensions (e.g. Virtual Worlds along technological quality, identification and likelihood). We argue that the former games tend to provide the player with a "complete" experience, thus should be more interesting in a medium-long term and for a larger audience. The latter typologies, on the other hand, should be particularly appealing to specialized and keen players. Understanding game typology characterization should have a significant relevance for the design of new games. In particular, we expect that successful games should have a similar characterization as their typology. This means that they should address in particular those player experience dimensions that are perceived as relevant for that specific typology. However, this is a speculation that has to be investigated with further testing. Finally, we have also tried to see if new, "hybrid" game typologies could be devised based on the combination of important constructs.

References

1. Barnes, T., Encarnação, L.M., Shaw, C.D.: Serious Games. IEEE Computer Graphics and Applications 29(2), 18–19 (2009)
2. Pine II, B.J., Gilmore, J.H.: Welcome to the Experience Economy. Harvard Business Review, 97 (July-August 1998)
3. Sliney, A., Murphy, D.: JDoc: A Serious Game for Medical Learning. In: Proc. International Conference on Advances in Computer-Human Interaction, pp. 131–136 (2008)
4. Bernhaupt, R., Eckschlager, M., Tscheligi, M.: Methods for evaluating games: how to measure usability and user experience in games? In: Proc. of the international Conference on Advances in Computer Entertainment Technology, Salzburg, Austria, June 13-15 (2007)
5. Bannister, D., Fransella, F.: Inquiring man, 3rd edn. Routledge, London (1985)
6. Hudson, R.: Images of the Retailing Environment: An Example of the Use of the Repertory Grid Methodology. Environment and Behavior 6, 470–494 (1974)
7. Hassenzahl, M., Trautmann, A.: Analysis of web sites with the repertory grid technique. In: CHI 2001 Extended Abstracts on Human Factors in Computing Systems, Seattle, Washington, March 31-April 5 (2001)
8. Mandryk, R.L., Atkins, M.S., Inkpen, K.M.: A continuous and objective evaluation of emotional experience with interactive play environments. In: Proc. of the SIGCHI Conference on Human Factors in Computing Systems, Montréal, Québec, Canada, April 22-27 (2006)
9. Kelly, G.: The psychology of personal constructs, vol. 1 & 2. Routledge, London (1955)
10. Tan, F.B., Hunter, G.M.: The Repertory Grid technique: A Method For The Study of Cognition in Information Systems. MIS Quarterly 1(26), 39–57
11. Federoff, M.: Heuristics and usability guidelines for the creation and evaluation of fun in video games. Unpublished thesis, Indiana Univ., Bloomington
12. Sweetser, P., Wyeth, P.: GameFlow: a model for evaluating player enjoyment in games. Computers in Entertainment 3(3) (July 2005)
13. Raney, A.: Expanding disposition theory: Reconsidering character liking, moral evaluations, and enjoyment. Communication Theory 14(4), 348–369 (2004)

14. Nabi, R., Krcmar, M.: Conceptualising media enjoyment as attitude: implications for mass media effects research. Communication Theory 4(14), 288–310 (2004)
15. Green, M., Brock, T., Kaufman, G.: Understanding media enjoyment: The role of transportation into narrative worlds. Communication Theory 14(4), 311–327 (2004)
16. Chen, J.: Flow in games (and everything else). Commun. ACM 50(4), 31–34 (2007)
17. Ranganathan, C., Ganapathy, S.: Key dimensions of business-to-consumer websites. Information & Management 39, 457–465
18. Clarke, D., Duimering, P.R.: How computer gamers experience the game situation: a behavioral study. Comput. Entertain. 4(3), 6 (2006)
19. Steed, A., McDonnell, J.: Experiences with Repertory Grid Analysis for Investigating Effectiveness of Virtual Environments. In: PRESENCE 2003, October 6-8. Aalborg Univ. (2003)
20. Boyd, S., Carini, D., Carini, C.: Constructing a Player-centered definition of fun for video games design. In: Proceedings of HCI 2004, Leeds Metropolitan University, UK (September 2004)
21. Hazlett, R.L.: Measuring emotional valence during interactive experiences: boys at video game play. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Montréal, Québec, Canada, April 22-27 (2006)
22. Yannakaki, N., Lund, H.H.: Entertainment capture through heart rate activity in physical interactive playgrounds. User Modeling and User-Adapted Interaction 18(1) (2008)
23. Larsen, J.M.: Evaluating User Experience - how game reviewers do it, Workshop on Evaluating User Experiences In Games. In: CHI 2008 Conf., Florence, Italy, April 5-10 (2008)
24. Bellotti, F., Berta, R., De Gloria, A., Zappi, V.: Exploring Gaming Mechanisms to Enhance Knowledge Acquisition in Virtual Worlds. In: 3rd International Conference on Digital Interactive Media in Entertainment and Arts (DIMEA 2008), Athens, Greece (September 2008)
25. Cowley, B., Charles, D., Black, M., Hickey, R.: Toward an understanding of flow in video games. Comput. Entertain. 6(2), 1–27 (2008)
26. Clarke, D., Duimering, P.R.: How computer gamers experience the game situation: a behavioral study. Comput. Entertain. 4(3), 6 (2006)

A Plot-Manipulation Algebra to Support Digital Storytelling

Börje Karlsson, Simone D.J. Barbosa, Antonio L. Furtado, and Marco A. Casanova

Departamento de Informática, PUC-Rio
Marquês de São Vicente, 225 / 4º andar RDC
Gávea, Rio de Janeiro, RJ, Brazil, 22451-900
{borje,simone,furtado,casanova}@inf.puc-rio.br

Abstract. Plot composition is examined here at a logic design level, an intermediate stage that comes next to the conceptual level wherein the intended narrative genre is specified. An abstract data structure is proposed to represent plots, together with an algebra for manipulating the data structure. Our purpose is to adapt for narratives the strategy applied to databases by Codd's relational model. The basic operators of our Plot-Manipulation Algebra (PMA) are introduced in view of the four fundamental relations between events that we identified in a previous work. A logic programming prototype was implemented in order to run examples using the algebra.

Keywords: storytelling, narratology, plots, logic design, algebraic formalisms.

1 Introduction

Narratology studies distinguish three levels in literary composition: fabula, story and text [4]. Here, we stay at the *fabula* level, where the acting characters are introduced, as well as the narrative *plot*, consisting of a partially-ordered set of *events*.

At least four concerns are involved in plot composition:

- the plot must be formed by a *coherent sequence* of events;
- for each position in the sequence, several *alternative choice*s may apply;
- non-trivial interesting sequences must permit unexpected *shifts along the way*;
- one may need to *go down to details* to better visualize the events.

These concerns led us to identify, drawing on linguistic and semiotic work [22, 7, 9], four relations between events, namely *syntagmatic*, *paradigmatic*, *antithetic*, and *meronymic* relations [11]. It turns out that such relations hold as a consequence of the conventions regulating the chosen narrative *genre*. Therefore, a necessary preliminary step to plot composition is to provide a *conceptual specification* of the genre.

In our conceptual modelling approach, we focus on events that correspond to the execution of predefined *operations*, deliberately performed by the characters. Each operation is defined in terms of its pre-conditions and post-conditions, and their interplay is what induces the partial order requirements for the plots, and constitutes the basis for characterizing the presence of the four kinds of relations between events.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 132–144, 2009.

However, a conceptual specification is still too far removed from a concrete computerized system to support plot composition. An analogous problem was successfully faced by database researchers, and, as we shall indicate, their three-stage solution can be conveniently adapted for our purposes. The key idea is to provide a *logic design* stage, mediating between conceptual design and physical implementation. This was very effectively achieved by the Relational Model proposal, which introduced a *relational algebra* [12] to model table manipulation at the logic level.

Codd's *relational model* is widely recognized today as providing effective guidance for database design at the logical level. The abstract data type on which it is based is the *n-ary relation*, also known as (relational) *table*, defined as a subset of the Cartesian product of n data domains: $R \subseteq D1 \times D2 \times ... \times Dn$. To handle database structures, Codd proposed [12] a *relational algebra* sublanguage, whose operator set can be reduced, as can be easily demonstrated [25], to five primitives: *product*, *projection*, *union*, *selection* and *difference*. All operators have tables as operands and yield tables as their result. Product, union and difference are called binary operations (since they involve two operands), whereas projection and selection are unary operations (one operand). Codd's operators (with two additions to accommodate the meronymic dimension) will serve our purposes if we substitute our plot abstract data type for tuples and, consequently, sets of plots (to be called *libraries*) for the relational tables.

Accordingly, we propose here a logic design stage for plot composition, involving an abstract structure for plots and a Plot-Manipulation Algebra (PMA) to handle the structure, taking into due account the relations between narrative events implied by the conceptual level specification of the genre. To illustrate the discussion, as well as the design and use of a logic programming prototype tool, we employ an example involving a small number of events, which, in different combinations, have been treated repeatedly in literary works.

A very simple PMA prototype was implemented to experiment with the notions discussed here. It serves to compose plots by applying the repertoire of algebraic operations, optionally resorting to pre-defined plots and plot libraries to supply useful clues. The entire system was written in SWI-Prolog. The examples illustrated throughout the paper were all executed in the PMA prototype.

The paper is organized as follows. Section 2 covers the background for our work. Section 3 presents the basic algebraic operators, illustrated by examples, a few extensions being added in section 4. Section 5 contains concluding remarks.

2 Basic Notions

2.1 Relations between Events

To illustrate the event relations, we shall employ a simple example to be referenced along the paper. Consider four types of events, all having one woman and two men as protagonists: *abduction*, *elopement*, *rescue*, and *capture*. As demonstrated in folktale studies [20], many plots mainly consist of an act of villainy, i.e. of a violent action that breaks the initially stable and peaceful state of affairs, followed ultimately by an action of retaliation, which may or may not lead to a happy outcome.

Syntagmatic Relations. To declare that it is legitimate to continue a plot containing abduction by placing rescue next to it, we say that these two events are connected by a syntagmatic relation. More precisely, the occurrence of the first leaves the world in a state wherein the occurrence of the second is coherent. Similarly, a plot involving elopement followed by capture looks natural, and hence these two events are likewise related. The syntagmatic relation between events induces a weak form of causality or enablement, which justifies their *sequential ordering* inside the plot.

Paradigmatic Relations. The events of abduction and elopement can be seen as *alternative* ways to accomplish a similar kind of villainy, and therefore there is a *paradigmatic relation* between them. Both achieve approximately the same effect: one man takes away a woman from where she is and starts to live in her company at some other place. There are differences, of course, since the woman's behaviour is normally said to be coerced in the case of abduction, but quite voluntary in the case of elopement. The same type of relation is perceived to hold between the events of rescue and capture, which are alternative forms of retaliation.

As the present example suggests, the so-called syntagmatic and the paradigmatic axes [22] are really *not* orthogonal in that the two relations cannot be considered independently when composing a plot. Thus, in principle, the two pairs abduction-rescue and elopement-capture are the only normal ones, as illustrated in several literary works (for specific examples, cf. [18]). Yet the next type of relation shows that such limitations can, and even should, be waived occasionally.

Antithetic Relations. While normal plots can be composed exclusively on the basis of the two preceding relations, the possibility to introduce unexpected turns is often desirable in order to make the plots more attractive (cf. the notions of *recognition* and *reversal* in [2]) – and this requires the construct that we chose to call *antithetic relation*. A context where a woman suffers abduction by a ravisher whom she does not love would seem incompatible with a capture event. So, in this sense, abduction and capture are in antithetic relation. However, if the woman has a change of heart and falls in love with the villain (perhaps as a case of *Stockholm syndrome*), she will refuse to be rescued, and will only return to her previous lover if captured.

Generally speaking, if some *binary opposition* (e.g. “to love or not to love” dilemma) is allowed to be manipulated via some agency external to the predefined events, then one can have plots that no longer look conventional. A sort of discontinuity is produced by such radical shifts in the context. Intervening between abduction and capture, or between elopement and rescue, a sudden change of feelings can give rise to these surprising sequences. Also, a change of beliefs may cause a reversal in the course of actions, usually in a totally opposite direction.

Figure 1 shows the relations thus far discussed.

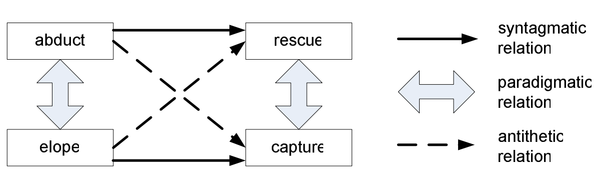


Fig. 1. Syntagmatic, paradigmatic, and antithetic relations

Meronymic Relations. *Meronymy* is a word of Greek origin, used in linguistics to refer to the decomposition of a whole into its constituent parts. Forming an adjective from this noun, we call *meronymic relations* those that hold between an event and a lower-level set of events, with whose help it is possible to provide a more detailed account of the action on hand.

Thus, we could describe the abduction of a woman called Sita by a man called Ravana (characters taken from the *Ramayana* [26]) as: "Ravana rides from Lanka to forest. Ravana seizes Sita. Ravana carries Sita to Lanka." And her rescue by Rama could take the form: "Rama rides from palace to Lanka. Rama defeats Ravana. Rama entreats Sita. Rama carries Sita to palace." (Figure 2).

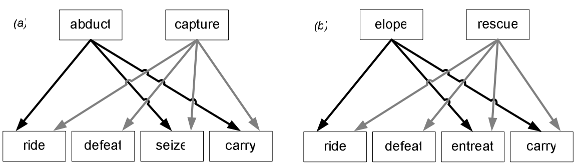


Fig. 2. Meronymic relations: (a) the forceful actions and (b) the gentle actions

Detailing is most useful to pass from a somewhat abstract view of the plot to one, at a more concrete physical level [23], that is amenable (possibly after further decomposition stages) to the production of a computer graphics animation [10]. Mixed plots, combining events of different levels, also make sense, satisfying the option to represent some events more compactly while showing others in detail.

2.2 Conceptual Specification of Genres

In order to model a chosen genre, to which the plots to be composed should belong, one must specify at least: (a) what can exist at some state of the underlying mini-world; (b) how states can be changed; and (c) the factors driving the characters to act.

Accordingly, we start with a conceptual design method involving three schemas – static, dynamic and behavioural – which has been developed for modelling literary genres encompassing narratives with a high degree of regularity, such as fairy tales, and application domains of business information systems, such as banking, which are constrained by a basically inflexible set of operations and, generally, by following strict and explicitly formulated rules [17]. Indeed, in our model, we equate the notion of event with the state change resulting from the execution of a predefined operation.

The *static schema* specifies, in terms of the *Entity-Relationship* model [5], the entity and relationship classes and their attributes (cf. the notion of story databases in [19]). In our simple example, character and place are entities. The attributes of characters are `name`, which serves as identifier, and `gender`. Places have only one identifying attribute, `pname`. Characters are pair-wise related by relationships `loves`, `held_by` and `consents_with`. The last two can only hold between a female and a male character; thus `held_by(Sita,Ravana)` is a *fact* meaning that Sita is forcefully constrained by Ravana, whereas `consents_with(Sita,Ravana)` would indicate that Sita has voluntarily accepted Ravana's proposals. Two relationships associate characters with places: `home` and `current_place`. A state of the world consists of all facts about the existing entity instances and their properties holding at some instant.

The *dynamic schema* defines a fixed repertoire of operations for consistently performing state changes. The *STRIPS* [16] model is used. Each operation is defined in terms of pre-conditions, which consist of conjunctions of positive and negative literals, and any number of post-conditions, consisting of facts to be asserted or retracted as the effect of executing the operation. Instances of facts such as `home` and `gender`, are fixed, not being affected by any operation. Of special interest are the *user-controlled* facts which, though immune to operations, could be, as suggested in [11], manipulated through arbitrary *directives*. In our example, `loves` is user controlled.

For the present example, we have provided operations at only two levels. The four main events are performed by level-1 operations: `abduct`, `elope`, `rescue` and `capture`. Operations at level-2 are actions of smaller granularity, in terms of which the level-1 operations can be detailed: `ride`, `entreat`, `seize`, `defeat`, and `carry`.

Our provisional version of the *behavioural schema* consists of *goal-inference* (a.k.a. *situation-objective*) *rules*, *belief rules*, and *emotional condition rules*. For the example, three goal-inference rules are supplied. The first one refers to the ravisher: when the princess is not at her home and the hero is not in her company – and hence she is unprotected – the ravisher will want to do whatever is adequate to bring her to his home. The other goal-inference rules refer to the hero when the ravisher has the woman in his home: either the hero believes that she does not love the other man, or he believes that she does. In both situations, he will want to bring her back, freely in the first case (`rescued`) and constrained in the second (`captured`).

Informally, beliefs correspond to the partial view, not necessarily correct, that a character currently forms about the factual context (for a formal characterization, cf. the BDI model [13, 21]). The belief rules that we formulated for our example look rational, but notice that they are treated as defaults, which could be overruled by a directive. A man (the hero or the ravisher) believes that the woman does *not* love his rival if the latter has her confined, but if she has ever been observed in his company and in no occasion (state) was physically constrained, the conclusion will be that she is consenting (an attitude seemingly too subjective to be ascertained directly in a real context). The emotional condition rules refer to the three characters. A man (or woman) is happy if currently in the beloved's company, and bored otherwise.

2.3 The Plot Abstract Data Type

A *plot* P is a pair [S,D], where: S, the *event-set*, is a set of tagged events, and D, the *dependency-set*, is a set of order dependencies, expressed as tag-pairs. Tags are terms of the form fi, where i is a positive integer. By convention, the tags in an event-set S are numbered consecutively, starting with f1, and the tagged events in S are required to be placed in some sequence compatible with the partial order requirements expressed in D. The order dependencies are determined exclusively on the basis of the satisfaction of post-conditions by pre-conditions, and any dependencies deducible by transitivity are omitted. Such conventions lead in general to simpler and more efficient algorithms to handle the data structure. As an example, consider:

```
P = [[f1:  ride(Ravana, Lanka, forest), f2: entreat(Ravana, Sita),
     f3: seize(Ravana, Sita))], [f1-f2,f1-f3]]
```

We say that plots Pi and Pj are *similar* if, even with events with different parameter values and placed in a different sequence in the event-set, they have the same:

- number and type of events
- order dependencies
- co-designation/ non-co-designation schemes

Co-designation (or, respectively, non-co-designation) allows (forbids) the occurrence of the same value at different parameter positions. Note, for instance, that in the example above Sita occurs in the second position of `entreat` and of `seize`; a plot with Guinevere in both places would meet the same co-designation requirement. To check if the order dependencies are the same, one looks for a renaming of the tags of one of the plots that renders the sets of order dependencies in both plots identical. Similar plots are *equal* if the values at the corresponding parameter positions are identical.

As any abstract data type, plots are equipped with *operators* to handle plots along the dimensions induced by the event relations: 1. syntagmatic: product and projection; 2. paradigmatic: union and selection; 3. antithetic: difference; 4. meronymic: factoring and combination. Unary operators have higher precedence, and binary operators obey the set-theory norm. Operands are plots or libraries represented explicitly or, if predefined, by their names. As a shorthand, single-event plots admit event syntax, e.g. `[[f1:ride(Ravana,Lanka,forest)],[]]` denoted by `ride(Ravana,Lanka,forest)`.

3 Basic Algebraic Operators

Product. Given two plots P1 = [S1,D1] and P2 = [S2,D2], their product P := P1 * P2 is a plot P = [S,D], where S contains the S1 and S2 events, and D has the dependency-pairs duly recomputed between the S events, regardless of their origin from S1 or S2. If one or both operands are (non-empty) libraries rather than plots, the result is a library containing the product of each plot taken from the first operand with each plot from the second, according to the standard Cartesian product definition. If one of the operands is the empty plot, denoted by [], the result of the product operation is the other operand, and thus [] behaves as the neutral element for product. The case of an empty library, rather than an empty plot, demanded an implementation decision; by analogy with the zero element in the algebra of numbers, it would be justifiable to determine that a failure should result whenever an empty library occurred as one or both operands. However we decided to return the other operand as result, i.e. to regard this case as a frustrated attempt to extend plots instead of an error. This option is consistent with our decision to denote both the empty plot and the empty list by [].

Example 1. The product `P1*P2` in plot format and template-driven natural language.

```
:- P1:= carry('Ravana','Sita','Lanka') * ride('Rama', palace,'Lanka') *
     defeat('Rama', 'Ravana'),
   P2:= ride('Ravana', 'Lanka', forest) * seize('Rama','Sita') *
     carry('Rama','Sita',palace),
   P := P1 * P2.
P = [[f1:ride(Ravana, Lanka, forest), f2:carry(Ravana, Sita, Lanka),
      f3:ride(Rama, palace, Lanka), f4:defeat(Rama, Ravana),
```

```
    f5:seize(Rama, Sita), f6:carry(Rama, Sita, palace)],
    [f1-f2, f2-f3, f3-f4, f4-f5, f5-f6]]
```

Ravana rides from Lanka to forest. Ravana carries Sita to Lanka. Rama rides from palace to Lanka. Rama defeats Ravana. Rama seizes Sita. Rama carries Sita to palace.

Projection. Given a plot P' = [S',D'], its projection P := proj [T] @ P' is a plot P = [S,D], where S only contains the events of S' specified in the projection-template T, ordered according to the position of the terms in T, and D only contains dependency pairs involving events placed in S. The projection-template T is a sequence of terms F:O, where F is a tag and O an event. F and O can be variables; if O is not a variable, some or all of its parameters can be variables. The S events receive new consecutive tags, starting from f1. The main use of projection is to extract passages.

If the operand is a library, the result is a library containing the projection of the plots of the operand library. Note however that, since libraries are sets, they cannot contain duplicates, which may arise as the consequence of a projection that suppresses events distinguishing two or more plots – and such duplicates are accordingly eliminated from the result. If the projection fails for some reason, e.g. because the projection-template T referred to a tag or event that did not figure in S', the result will be the empty plot (or empty list) [] rather than an error.

Example 2. Projection can be simply used to re-order the events in a plot, in which case care must be taken to ensure that the new sequence be also viable. Given a predefined fake abduction plot, a new plot P can be obtained, preserving the original events but inverting the position of the third and fourth events. Instead of the predefined fake abduction plot, we now have a situation in which the villain initially acts as a seducer but, after bringing the princess to his home, decides to confine her.

```
fake_abduct(V,W,F) :- F := [[f1:ride(V,P1,P2),f2:entreat(V,W),
   f3:seize(V,W),f4:carry(V,W,P1)],[f1-f2,f1-f3,f2-f4]].
:- P := proj [f1:_,f2:_,f4:_,f3:_] @
fake_abduct('Ravana','Sita').
P = [[f1:ride(Ravana, Lanka, forest), f2:entreat(Ravana, Sita),
      f3:carry(Ravana, Sita, Lanka), f4:seize(Ravana, Sita)],
     [f1-f2, f2-f3, f3-f4]]
```

Ravana rides from Lanka to forest. Ravana entreats Sita. Ravana carries Sita to Lanka. Ravana seizes Sita.

Union. Given two operands U1 and U2, each of them either a plot or a library, their union U := U1 + U2 will always be a library containing all plots in U1 and U2, no two equal plots being retained. One or both operands can be the empty library, ambiguously denoted as said before by [], naturally functioning as the neutral element for union. If one of the operands is a plot and the other is [], their union is a library consisting of this single plot.

Selection. Given a plot P' = [S',D'], its selection P := sel [T]/E @ P' is the plot P' itself if the matching of the selection-template T against P' succeeds, as well as the subsequent evaluation of the logical expression E (whose presence is optional, except if T is empty), also involving information taken from P'. If the test fails, the result to be assigned to P is the empty library []. Usually the operand of selection will be a library, resulting in a library containing all plots that satisfy the test, or the empty library [] if none does. In order to select one plot at a time from a library L, even if L just contains a single plot, the form P := sel [T]/E @ one(L) should be employed.

Example 3. Consider a plot P that starts with abduction, and ought to continue with an adequate form of retaliation. Choosing this second event can be done by performing a (conditional) union of the two possible alternatives, making the effective presence of each of them depend on the recurring dilemma: the hero can only rescue the woman if she loves him, otherwise he must capture her. Notice that this scheme works as an exclusive-or or as an if-then-else. Assume that Sita currently loves Rama.

```
:- P := abduct(V,W) *
     ((sel []/loves(W,H) @ rescue(H,W)) +
      (sel []/(not loves(W,H)) @ capture(H,W))).
P = [[[f1:abduct(Ravana, Sita), f2:rescue(Rama, Sita)], [f1-f2]]]
```

Ravana abducts Sita. Rama rescues Sita.

Difference. Given two operands U1 and U2, each of them either a plot or a library, their difference U := U1 – U2 will always be a library containing all plots in U1 that are not equal to any plot in U2. As happens with standard set difference, the result of the operation is not affected by plots in U2 that have no equals in U1. On the other hand if all plots in U1 have their equals in U2, the empty list [] is assigned to U.

Example 4. The evaluation of plot P below illustrates the interplay of the syntagmatic, paradigmatic and antithetic event relations. The union operation yields a pair of libraries, consisting of the alternatives, respectively, for villainy and for retaliation. All villainy-retaliation sequences are then formed by the product of the two libraries. Finally, by difference, the transgressive sequence elope-rescue is excluded.

```
:- P := (abduct(V,W) + elope(V,W)) * (rescue(H,W) + capture(H,W)) -
        elope(V,W) * rescue(H,W).
P = [[[f1:abduct(Ravana, Sita), f2:rescue(Rama, Sita)], [f1-f2]],
     [[f1:abduct(Ravana, Sita), f2:capture(Rama, Sita)], [f1-f2]],
     [[f1:elope(Ravana, Sita), f2:capture(Rama, Sita)], [f1-f2]]]
```

Ravana abducts Sita. Rama rescues Sita.
Ravana abducts Sita. Rama captures Sita.
Ravana elopes with Sita. Rama captures Sita.

Factoring. Given a plot P' = [S',D'], its factoring P := fac P' is a plot P = [S,D], where each level-1 event ei' present in S' is replaced by a sequence ei1, ei2,..., ein of level-2 events. Each resulting sequence is obtained from a predefined map declaration map(Ei',[Ei1,Ei2,...,Ein]), such that Ei' matches ei'. In map declarations, all terms in both arguments are events, and may contain variables at the parameter positions. The first argument must be a level-1 event and the second a sequence of level-2 events.

When specifying a map declaration for an event Ei', care should be taken that the indicated sequence of level-2 operations should work as a plan successfully applicable at world situations satisfying the pre-conditions of Ei', and producing at the end the effects expressed by the post-conditions of Ei'. And if the sequence may have other (secondary) effects, these must not contradict those expected from executing Ei'.

As a map declaration for Ei' is found to match an event ei' in the course of factoring, all variables in Ei' will be instantiated with the values contained in the respective parameter positions of ei', with the consequence that several parameters of the level-2 events will also be instantiated by consistent variable substitution. In most cases, however, several level-2 events will not become fully ground terms. In order to further instantiate the level-2 event parameters, we decided to apply a heuristic process based on the pre-condition declarations of these events. As said before, some database facts of the mini-world may be invariant, in the specific sense that none of the events provided may change them, an example being the `gender` of the acting characters. So, if e.g. there is only one female character, and the pre-condition of an event requires a female at a certain parameter position, it seems natural to assign that character's name to the corresponding variable. If the operand is a library, the result is a library with the factoring of all the operand's plots.

Combination. Given a plot P' = [S',D'], its combination P := comb P' is a plot P = [S,D], where each sequence ei1', ei2',...,ein' of level-2 events present in S' is replaced by a level-1 event obtained by the inverse application of the pre-defined map declaration whose second argument happens to match the sequence. If the operand is a library, the result is a library containing all plots in the operand library with the eventual modifications performed by applying the combination operation.

Example 5. Using selection over the detailed description of the two level-1 villainy events, obtained through the factoring operation, we may determine which one involves the violent `seize` level-2 event. The desired level-1 event is reconstituted from the selected level-2 description by applying the combination operator at the end.

```
:- P := comb sel [_:seize(_,_)] @ (fac (abduct(V,W) + elope(V,W))).
P = [[[f1:abduct(Ravana, Sita)], []]]
```

Ravana abducts Sita.

4 Extensions

Two features extend the product operator, to achieve repeated plot sequences.

Given a plot P' = [S',D'], the n^{th} power of P', expressed by P := P' ** N, for a non-negative integer N, is evaluated according to the recursive formula:

- `if N = 0, P = []`
- `else P = P' * (P' ** (N - 1))`

Given a plot P' = [S',D'], the iteration of P', expressed by P := E@P', where E is a logical expression sharing any number of variables with P', is evaluated as follows:

- first, the iterator-template T is obtained, as the set of all possible instantiations of E at the initial state, and then:
- `if T is {}, P = []`
- else, if T = {t1, t2, ..., tn}, P = P'_{t1} * $P'_{\{t2, ..., tn\}}$

where P'_{ti} denotes P' with its variables instantiated consistently with those in ti, and the subscrit in $P'_{\{ti+1, ..., tn\}}$ refers to the remaining instantiations of T to be used at the next stages. As with product, both features apply to single plots and plot libraries.

Example 6. Sequences of actions with the same objective are often repeatedly played by different characters, especially in folktales [20]. Here, iteration is applied to make the two men successively attempt to persuade Sita, taking her feelings with respect to them into consideration when deciding how to act. To a man loved by Sita it is enough to entreat her once, whereas an unloved man might try that twice, finally resorting to seizing the reluctant princess.

```
:- P := iter (gender(M,male),home(M,H),gender(W,female)) @
     ( (sel []/(loves(W,M))@
        (ride(M,H,forest) * entreat(M,W))) +
       (sel []/(not loves(W,M)) @
        (ride(M,H,forest) * (entreat(M,W)**2 * seize(M,W)))) ).
```

Suppose Sita loves the hero, Rama, but not the villain, Ravana. Then the logical expression controlling the iteration operator will evaluate to the iterator-template T as shown below. Note that T is represented as a list with two items indicating, as required, different instantiations for the variables.

```
T = [ (M=Rama, H=palace, W=Sita),
      (M=Ravana, H=Lanka, W=Sita) ]
P = [[f1:ride(Rama, palace, forest), f2:entreat(Rama, Sita),
      f3:ride(Ravana, Lanka, forest), f4:entreat(Ravana, Sita),
      f5:entreat(Ravana, Sita), f6:seize(Ravana,Sita)],
     [f1-f2, f3-f4, f3-f5, f4-f6, f5-f6]]
```

Rama rides from palace to forest. Rama entreats Sita. Ravana rides from Lanka to forest. Ravana entreats Sita. Ravana entreats Sita. Ravana seizes Sita.

The fact that the current version of PMA is embedded in logic programming facilitated the introduction of *plot patterns*, allowing selection from libraries whose plots have events pertaining to the genre on hand, but are defined over different parameter values (e.g. names of persons, places, etc.). This is a first step towards the reuse of repositories of typical narratives collected from diverse sources (for folktale narratives, see for instance [1]), so as to exploit imitation as a composition resource.

Given a plot P = [S,D], the pattern *P* of P, expressed by *P* := patt P, is obtained from P by substituting variables for the parameters and tags in both S and D. As always, patterning applies to both single plots and to entire libraries. To take advantage of this feature, we provided an additional version of the selection operator, in which a plot can be used as selection-template. Let P' = [S',D'] be a plot at an early phase of composition, still with very few events, and let L be a library with the characteristics above. If one wishes to extend P' to a set of fuller alternative plots containing all events in S' and preserving the order requirements imposed by D', it is possible to adapt the typical narratives in L by way of a pattern-matching technique. This is accomplished by evaluating the expression P := sel P'@ (patt L), wherein plot P' guides a selection against the result of converting all plots in L into patterns.

Example 7. Library `lib_2`, displayed below, is used to extend an initial plot `Pi`.

```
lib_2([ [[f1:ride('Meleagant','Gore',forest),
          f2:seize('Meleagant','Guinevere'),
          f3:carry('Meleagant','Guinevere','Gore'),
          f4:ride('Lancelot','Camelot','Gore'),
```

```
            f5:defeat('Lancelot','Meleagant'),
            f6:entreat('Lancelot','Guinevere'),
            f7:carry('Lancelot','Guinevere','Camelot')],
            [f1-f2,f2-f3,f3-f4,f4-f5,f5-f6,f6-f7]],
          [[f1:ride('Tristan',forest,garden),
            f2:entreat('Tristan','Isolde'),
            f3:carry('Tristan','Isolde',forest),
            f4:ride('Mark','Cornwall',forest),
            f5:defeat('Mark','Tristan'),
            f6:seize('Mark','Isolde'),
            f7:carry('Mark','Isolde','Cornwall')],
          [f1-f2,f2-f3,f3-f4,f4-f5,f5-f6,f6-f7]] ]).
```

If `Pi` is formulated as shown next, the Meleagant plot will serve as model to obtain the extended plot `Pe`. Notice the later application of combination to `Pe` for recognizing the kind of narrative obtained, which turns out to be of the abduct-rescue variety.

```
:- Pi = [f1:carry('Ravana','Sita','Lanka'),f2:entreat('Rama','Sita'),
           f3:carry('Rama','Sita',palace)],
          [f1-f2,f2-f3]],
```

Ravana carries Sita to Lanka. Rama entreats Sita. Rama carries Sita to palace.

```
:- Pe := sel Pi @ (patt lib_2),
:- Pc := comb Pe.
Pe = [[[f1:ride(Ravana, Lanka, forest), f2:seize(Ravana, Sita),
        f3:carry(Ravana, Sita, Lanka), f4:ride(Rama, palace, Lanka),
         f5:defeat(Rama, Ravana), f6:entreat(Rama, Sita),
         f7:carry(Rama, Sita, palace)],
        [f1-f2, f2-f3, f3-f4, f4-f5, f5-f6, f6-f7]]]
```

Ravana rides from Lanka to forest. Ravana seizes Sita. Ravana carries Sita to Lanka. Rama rides from palace to Lanka. Rama defeats Ravana. Rama entreats Sita. Rama carries Sita to palace.

```
Pc = [[[f1:abduct(Ravana, Sita), f2:rescue(Rama, Sita)], [f1-f2]]]
```

Ravana abducts Sita. Rama rescues Sita.

A different result would be obtained by applying, to the same library, a different initial plot `Pi` diverging from the previous formulation with respect to the second event (`seize`, instead of `entreat`). With this the Tristan plot would be taken as model, resulting in a plot of the elope-capture variety.

5 Concluding Remarks

The contribution expected from a logic model is to provide reliable guidelines to develop systems that may be regarded as "complete" according to some criterion. We claim that PMA is complete in the specific sense that it covers plot manipulation along the dimensions induced by the syntagmatic, paradigmatic, antithetic and meronymic event relations. These relations encompass some fundamental aspects of plot composition, and are respectively associated, as argued in [11], with the *four major tropes* (metonymy, metaphor, irony, and synecdoche) of semiotic research [7, 9].

Of primary significance is the syntagmatic axis, along which plots are created by the *product* operator, whereby events or event-sequences are chained together to form progressively longer sequences. Inversely, a passage of interest can be extracted from a plot by the *projection* operator, possibly re-ordering the events while preserving the partial order. Notice in particular that the familiar cut-and-paste tactic corresponds to projections followed by product. Along the paradigmatic axis, the *union* operator offers different alternatives for certain positions in a sequence. Libraries, as sets of optional versions of narratives, are collected and expanded by applying union. Choosing among alternatives is the objective of the *selection* operator. It allows to check if a given plot has (or which plots in a library have) the desired characteristics.

The antithetic constraints, originating from binary oppositions in the mini-world context [9], require the *difference* operator, whose basic purpose is to exclude unwanted plots from a library collection. Curiously, what is negated once may, on second thought, suggest exciting directions for continuing the narrative, as posited in literary studies on the irony trope [6] and on the notion of *deconstruction* [14]. Such unexpected turns usually require that current facts and beliefs be changed, so that the the next events may seem to *blend* [15] naturally with the sequence so far composed. Finally, the concern with the meronymic dimension led to the inclusion of the *factoring* and *combination* operators, allowing, respectively, to detail or summarize the events. The ability to tune the level of granularity of events is vital, on the other hand, to prepare for the transition to the subsequent stages of story and text [4].

Theoretic work is needed to systematically investigate the formal properties of the algebra. Practical work, drawing from our early experiments with the **PlotBoard** prototype [11] and from differently oriented implementations [3,8,24], will include the design of systems based on PMA, taking maximum advantage of the previously developed plan-generation facilities, but also allowing effective user interaction along a step-wise plot composition and adaptation process through a friendly interface. Any such system should have access to an online representation of the conceptual schemas, using this meta-level information to keep checking the *semantic correction* of the plots being generated. Moreover, access to the behavioural schema in particular should serve to verify what might be called *pragmatic plausibility*, i.e. whether the events caused by each character reflect their expected way of reacting, especially in view of the declared situation-goal rules.

References

1. Aarne, A., Thompson, S.: The Types of the Folktale. Suomalainen Tiedeakatemia (1987)
2. Aristotle. “Poetics”. In: Classical Literary Criticism. P. Murray & al (tr.). Penguin (2000)
3. Aylett, R., Louchart, S., Dias, J., Paiva, A., Vala, M., Woods, S., Hall, L.: Unscripted narrative for affectively driven characters. IEEE Computer Graphics and Applications 26(4), 42–52 (2006)
4. Bal, M.: Narratology. U. of Toronto Press (2002)
5. Batini, C., Ceri, S., Navathe, S.: Conceptual Design. Benjamin Cummings (1992)
6. Booth, W.: A Rhetoric of Irony. U. of Chicago Press, Chicago (1974)
7. Burke, K.: A Grammar of Motives. U. of California Press, California (1969)
8. Cavazza, M., Charles, F., Mead, S.J.: Character-based interactive storytelling. IEEE Intelligent Systems 17(4), 17–24 (2002)
9. Chandler, D.: Semiotics: The Basics. Routledge (2007)

10. Ciarlini, A.E.M., Pozzer, C.T., Furtado, A.L., Feijó, B.: A logic-based tool for interactive generation and dramatization of stories. In: Advances in Computer Entertainment Technology (2005)
11. Ciarlini, A.E.M., Barbosa, S.D.J., Casanova, M.A., Furtado, A.L.: Event Relations in Plan-Based Plot Composition. ACM Computers in Entertainment (to appear, 2009)
12. Codd, E.F.: Relational completeness of data base sublanguages. In: Rustin, R. (ed.) Database Systems. Prentice-Hall, Englewood Cliffs (1972)
13. Cohen, P.R., Levesque, H.J.: Intention is Choice with Commitment. Artificial Intelligence 42 (1990)
14. Culler, J.: On Deconstruction. Cornell U. Press (1983)
15. Fauconnier, G., Turner, M.: The Way We Think. Basic Books, New York (2002)
16. Fikes, R.E., Nilsson, N.J.: STRIPS: A new approach to the application of theorem proving to problem solving. Artificial Intelligence 2 (1971)
17. Furtado, A.L., Casanova, M.A., Barbosa, S.D.J., Breitman, K.K.: Analysis and Reuse of Plots using Similarity and Analogy. In: Li, Q., Spaccapietra, S., Yu, E., Olivé, A. (eds.) ER 2008. LNCS, vol. 5231, pp. 355–368. Springer, Heidelberg (2008)
18. Karlsson, B., Furtado, A.L.: PMA: A Plot Manipulation Algebra. Technical Report MCC 01/09, PUC-Rio (2009), ftp://ftp.inf.puc-rio.br/pub/docs/techreports/09_01_karlsson.pdf
19. Oinonen, K., Theune, M., Nijholt, A., Uijlings, J.: Designing a Story Database for Use in Automatic Story Generation. In: Harper, R., Rauterberg, M., Combetto, M. (eds.) ICEC 2006. LNCS, vol. 4161, pp. 298–301. Springer, Heidelberg (2006)
20. Propp, V.: Morphology of the Folktale. S. Laurence (trans.). U. of Texas Press (1968)
21. Rao, A.S., Georgeff, M.P.: Modeling rational agents within a BDI-architecture. In: Int'l Conf. on Princ. of Knowledge Representation and Reasoning (1991)
22. Saussure, F.: Cours de Linguistique Générale. In: Bally, C., et al. (eds.) Payot (2006)
23. Schank, R.C., Colby, K.: Computer Models of Thought and Language. Freeman, New York (1973)
24. Szilas, N.: ID-tension: A narrative engine for interactive drama. In: Proc. of Technologies for Interactive Digital Storytelling and Entertainment (2003)
25. Ullman, J.D., Widom, J.: A first Course on Database Systems. Prentice-Hall, Englewood Cliffs (2008)
26. Valmiki: Le Ramayana. Benoît, P., et al. (trans.). Gallimard (1999)

Distributed Episode Control System for Interactive Narrative Entertainment

Jun'ichi Hoshino, Katsutoki Hamana, Shiratori Kazuto, and Atsushi Nakano

University of Tsukuba, Graduate school of Systems and Information Engineering,
1-1-1, Tennodai, Tsukuba-shi, Ibaraki, Japan
{hamana,nakano}@entcomp.esys.tsukuba.ac.jp,
jhoshino@esys.tsukuba.ac.jp

Abstract. We propose the massive action control system (MACS) for interactive narrative entertainment. MACS determines the action priorities for characters based in part on their own internal states, such as the motivation behind the action, feeling, and personality. MACS selects a behavior control module, called an episode tree, of about 1000 events, which is divided into action types based on these internal states and external situations. We demonstrate the effectiveness of the system with the Spilant World interactive animation contents at the National Museum of Emerging Science and Innovation in Japan, and NAMCO amusement park.

Keywords: narrative entertainment, massive action, episode tree, lifelike.

1 Introduction

In recent years, many entertainment systems have relied on the progress of interaction technology to create characters that act autonomously. To show these lifelike characters, it is important for them to perform various actions, such as daily actions, reflex actions that require reacting to input from a user, perceiving actions where the character perceives an object and reacts to it, and actions based on personalities or feelings. This results in the problem of complex action planning. A character has to carry out the actions listed, keep schedules and maintain a personality, and react flexibly to user interaction, while still maintaining story flow.

In this paper, we propose the massive action control system (MACS), which can execute various actions in multiple characters (Fig. 1). This system continuously selects an appropriate fragmentary behavior control module, called an episode tree, based on the character's inner states, such as motives, feelings, and personality, and the state of the external world, such as other characters and objects surrounding the character. Thus, the character can also respond to freely timed user interference. It has three main features:

- **Structuring Actions:** Conditional branching becomes very complex when the control system treats individual actions that are elements of units, such as standing, sitting, or waving. Thus, we define each action as one structure (episode tree), where multiple elements and the start or end conditions are combined hierarchically.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 145–156, 2009.

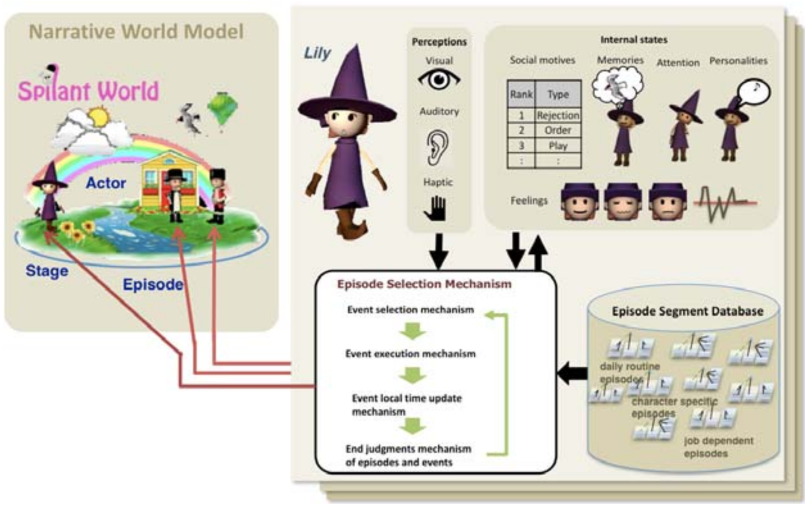


Fig. 1. MACS system configuration

Fig. 2. Spilant World narrative entertainment system at NAMCO amusement park. Operating everyday from September 2008.

- **Prioritizing Actions:** A character's intentions are important in entertainment systems. Therefore, our system prioritizes actions based on internal states, such as the character's personality, significant motivations behind the actions, and the character's feelings at that time.
- **Interpolating Actions:** If a character's action sequence changes without notice, the action will not seem natural. Therefore, it is necessary to create transitional actions to link a sequence of actions smoothly together. Pre-processing before each episode tree node assists in transitioning between actions and returning to an action performed in the past. Post-processing cleans out previous actions at the end of an episode and when current actions are interrupted. MACS interpolates the actions by inserting special processes automatically.

We created an interactive entertainment system called Spilant World (Fig. 2) to demonstrate the action control system. Multiple characters have motivations behind their actions, feelings, and personalities, and live daily in this application.

2 Related Work

One of the problems in story-based interactive entertainment is action planning. Characters have schedules such as work, meals, and physiological phenomena, as well as unique parameters such as memories and feelings, like humans in the real world. The SIMS 2, produced by Electronic Arts (EA), and Kenran Buto Sai, produced by Sony Computer Entertainment Inc. (SCEI), are typical examples of such games. Tale-Spin [13] generates a story from facts in the virtual world and rules governing the behavior of the characters. Improv [17] can produce an actor that initiates reactions in the user or other actors by a script in real time. The Oz-Project [11] proposes an interactive drama in which speech develops from interaction between the user and a globular character that has eyes, nose, and a mouth, and displays feelings. Other approaches include controlling characters' actions using the Cognitive Modeling Language (CML), which can intuitively impart knowledge to the character about an action and the preconditions influencing it [7], story generation with dynamic planning of affiliate of individual action in Dual Dijkstra's search for planning [14], and controlling characters' actions by describing a story arc that uses Hierarchical Task Networks (HTN), i.e., a task tree arranged hierarchically [4]. Façade is proposed as an interactive drama in which the game progresses by natural-language conversation as a more reformative work [12].

Furthermore, investigations of systems where the user can interact with animal characters have also been advanced. Live World [20] is a system where the user can make an object and give life to it, with the altered object behaving like an animal. Dobie T. Coyote [1] and Alpha Wolf [19] are applications that use reactions to interaction with the user. As for the former, the user can feel the vitality by seeing the dog character that learns in the user's training by the remote control training actually used. The latter is an application that reproduces social actions of a wolf; the user can see lifelike animation through reproduction of the wolf's reaction based on the actions of the wolfpack. The user can interfere in the wolf's activities by barking into the microphone.

3 Massive Action Control System

3.1 MACS Mechanisms

To construct lifelike animated characters, it is important to have them perform various actions, such as daily tasks, reflexes, acts based on their perceptions, and actions based on personalities or feelings. The MACS performs these various actions in complex narrative situations. Each character uses the system (Fig. 1). The inputs to MACS are sensory information from the external narrative world, the character's current actions, and its internal status. MACS outputs various actions based on the input information and the stored episode group. Episodes in which multiple characters interact are shared by the characters.

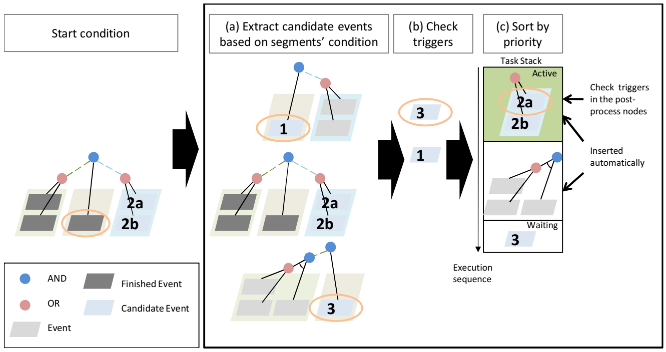


Fig. 3. Event selection mechanism

This system consists of four mechanisms of event selection, event execution, updating event time, and end judgments for episodes/events (Fig. 1).

- **Event Selection Mechanism:** This mechanism extracts events satisfying AND/OR conditions of a higher rank class from the episode tree under execution and selects an event having the highest priority in the episode. As illustrated in Fig. 3, the event selection process is as follows.
  1. Extracting candidates for the next event considering a recently finished event (a).
  2. Adding candidates to task lists if conditions in an event are satisfied (b).
  3. Sorting tasks in order of the priority (c).
  4. Selecting the event having the highest priority as the next event (c).
- **Event Execution Mechanism:** This mechanism performs all actions included in the selected event. An event has a local timeline, which begins when the event starts. The characters perform the appropriate action corresponding to a moment in local time.
- **Event Local Time Update Mechanism:** This mechanism updates local time in events.
- **End Judgment Mechanism for Episodes and Events:** This mechanism judges the end of episode trees and events.

By repeating and processing these mechanisms, each character selects and executes events. However, when multiple characters share the same talk like the narrative as an exception, these characters share the same event and perform it synchronously.

3.3 Action Structures and Prioritizing Actions

As described in Figure 4, an episode tree has AND/OR nodes in its higher rank class. The AND node becomes true after all the events of the child nodes are performed and finished. The OR node becomes true after at least one of the child nodes is performed and finished. The flow of an episode is made easy to see by expressing its abstract contents by the nodes in the higher rank class of its tree. The events are arranged in the lowermost part of the tree. An event is a structure with a trigger and an action, and

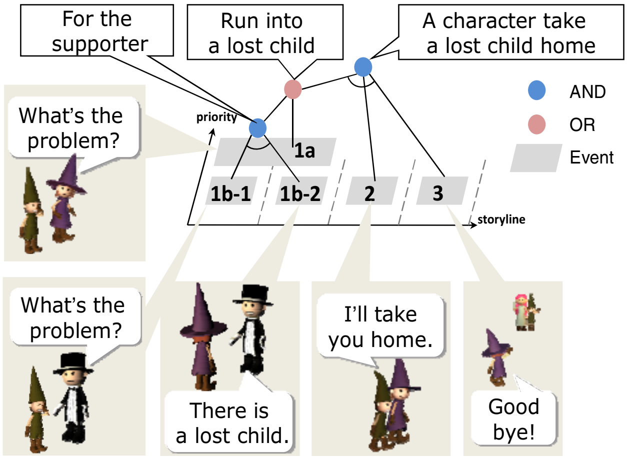


Fig. 4. Episode tree and Event contents

when the conditions used as a trigger are fulfilled, it performs the specified action. A rightward story line expresses the story's direction of movement, which advances by changing the events chosen as candidates, judging the AND/OR conditions specified as the node of the tree.

To create lifelike characters, the character's actions must be prioritized based on its personality. MACS assigns priority to actions based on Murray's 20 social action motives [15]. Each episode tree is a set of actions, and so it refers to one of these motives. For each character, MACS also ranks the motives to which that character should give priority, and compares the ranking of motives when executing an action. Even when the episode tree fulfills conditions based on situations in the existing external world. MACS selects an episode tree close to the action motive of a character using these priorities, and thus gives individuality to a character.

In the proposed system, reflex actions and daily actions are treated exceptionally in order to make a character lifelike. Reflex actions are performed without thinking, and their priority is always the highest. In this system, they are actions that occur when a character is touched or grasped. Daily actions show daily life. Even when the character perceives nothing, it is not lifelike unless it carries out its regular actions. We then prepare an episode tree of daily actions that have the lowest priority in the episode tree group, and even when not fulfilling any particular conditions, the character follows its routine.

3.4 Action Interpolation

This section describes an action interpolation function. The pre-processing and post-processing tasks corresponding to each event can be arranged as nodes of an episode tree. When moving from the current event to the next event, the action is appropriately interpolated by inserting the following processing.

1. Confirm whether the post-processing node has been placed at the node of the higher rank class of the event that has been selected now.

2. If so, add it to the task list as a top-priority task in post-processing.
3. Confirm whether the pre-processing node has been placed to a higher rank class of the selected event.
4. If so, add it to the task list as a top-priority task if the post-processing is not inserted as the task of executing it as follows if it has been inserted.
5. Execute tasks in the following order: the post-processing task, the pre-processing task added to the task list, and the event that is selected and executed next.

The following two cases can occur during the interpolation processing.

A) The event selected next changes when the post-processing task is executed.
B) The event selected next changes when the pre-processing task is executed.

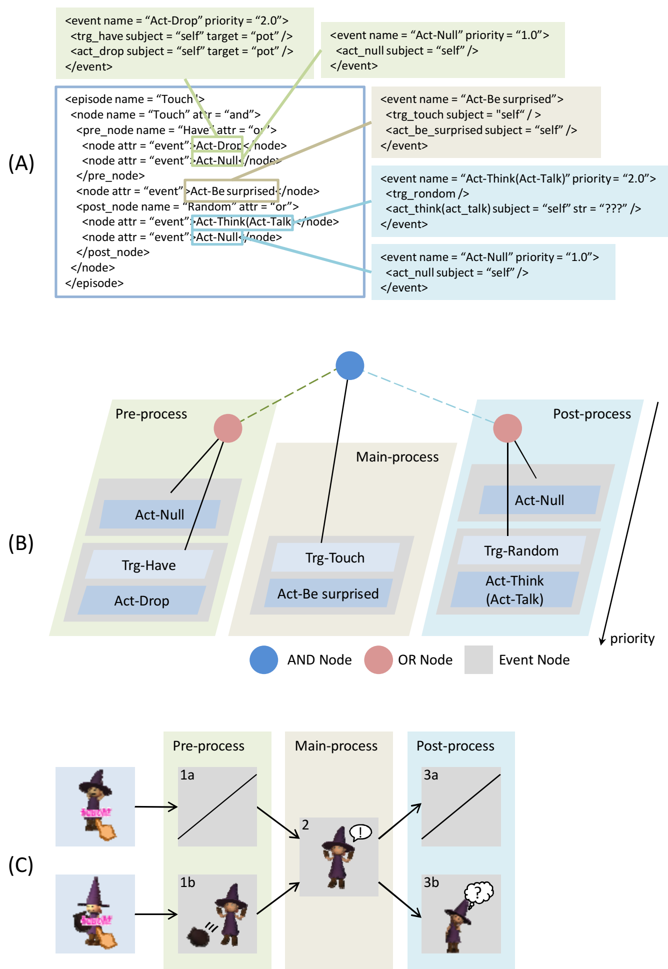


Fig. 5. (A) Scripting episode tree using XML. (B) Illustrating the script: episode tree. (C) Character's action by the script: reflex action.

The processing when these two cases arise is as follows. In Case A, all pre-processing tasks stacked to the task list are deleted, but the post-processing task is certainly executed because it is a resolution of the preceding event. Alternatively, if the pre-processing tasks exists in a higher rank class of the event that will be newly selected and executed next, they are added to the task list. In Case B, if the post-processing tasks exist in a higher rank class than the pre-processing task that will be executed next, they are added, and the action to schedule for execution in the interim is determined. In addition, if the pre-processing tasks exist in a higher rank class of the event that will be newly executed, they are added to the task list.

4 Spilant World Narrative Entertainment System

This section describes an interactive application called Spilant World in which a user can drag an icon in the narrative world and also appreciate the various character's reactions by touching or grasping the object in the virtual world. An early version of the application is presented in [16]. In this paper, we extend the Spilant World system using MACS, and demonstrate the effectiveness of the system at the National Museum of Emerging Science and Innovation in Japan, and NAMCO amusement park.

4.1 System Overview

This system consists of a 37' LCD equipped with an optical touch panel, speaker, and PC. A virtual world is displayed, and the user can participate in the world by placing a finger on the touch panel. The screen display is as follows. The virtual world has a vivid display with the icons for object addition located along the right-hand side. The user can perform four operations: touching an object or character with a finger; creating a new object by holding an icon and dragging and dropping it in the virtual world; grasping a character or object by touching for an extended time; and drawing a line by touching the screen and moving the hand quickly. When performing a direct interaction on a character via touching/grasping, the user can see the character's reaction. Moreover, the character's reaction can be seen when the user interacts indirectly by adding an object and drawing an event line. (See Fig. 6.)

4.2 Number of Episodes and Aliveness

We experimented following two cases and compared it. Case 1 is when the number of episode trees is 321 and Case 2 is when it is 946. However, in practice, since it is removed from the candidate for search by the motive of a character, the former is 100, and the latter is 390. Comparison of action of the character about those two cases is shown in Fig. 9. In (A), the user added a bloom object. In (B), the user grasped and moved a character. And in (C), the user touched a light. The horizontal axis is progress of time.

(A) In Case 1, a character reacted to a broom, said “It's a broom! I'll take it.”, and approached it. Then, she took it and only walked. And, another character didn't react to her with a broom. It compares, in Case 2, the occurrence until she speaks is the same as a Case 1, but, she flew using the broom. Then, another character was surprised at the sight.

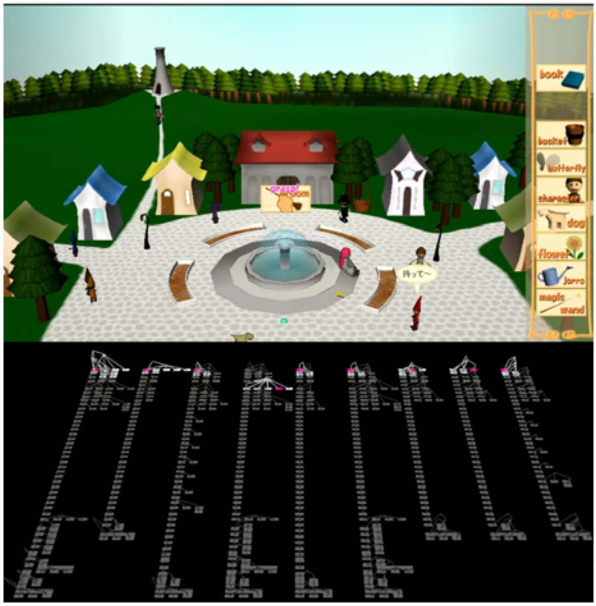


Fig. 6. A screen capture from Spilant World system with internal episode status. Active episode trees are highlighted. Red modules are currently executed by the characters.

(B) In Case 1, when the grasped character was moved, there was some characters were surprised it and there was also some characters were not surprised it. It compares, in Case 2, almost all the characters that are present in the neighborhood show the reaction.

(C) In Case 1 and in Case 2, the light had be lighting similarly first. Then, the blue character both indicated the reaction "It shines!" However, after that, in Case 1, only he showed interest and other characters had not looked at all. It compares, in Case 2, they said "Beautiful!" and "Oh!" Each character reacts to the light.

4.3 Demonstration

Fig. 7 illustrates the situation at the National Museum of Emerging Science and Innovation in Japan. Many children have experienced our system. Then, we recorded the user's operation and the state of event and episode selected by the character/object. The number of episode trees is 128 and the number of events is 434. The items for every object are as in Table 2.

4.3.1 Evaluation by Record

Record of the event/episode for every character/object is described. As an example, top 15 pieces are shown in Table 3 about the number of times of an event appearance of a character called "Lily". The event which consecutive numbers attach in the table is an event in the same episode, and shows the appearance order. As shown also in this table, schedule action of WalkAround (event which takes a walk) etc. has appeared mostly, and it is because the character was always acting along with the

Fig. 7. Many children have experience our Spilant World system at the National Museum of Emerging Science and Innovation in Japan

Table 2. The number of episode trees and events

Character	The number of episode trees	The number of events
Lily	25	107
Toto	21	77
John	21	67
Sala	24	80
Scot	7	22
Tom	4	13
Dog	13	29
Others	13	39
Total	128	434

Table 3. The number of times and the frequency of appearance of events

Order	Event	The number of times	The frequency
0	WalkAround1	548	14.96
1	WalkAround2	329	8.98
2	grasp_event1	234	6.39
3	grasp_event2	230	6.28
4	WalkAround3	119	3.25
5	WalkAround4	99	2.7
6	fall_fountain_event3	97	2.65
7	on_the_roof1	93	2.54
8	fall_fountain_event2	80	2.18
9	on_the_roof2	74	2.02
10	on_the_roof3	67	1.83
11	find_broom_pre	54	1.47
12	find_butterfly1	54	1.47
13	find_butterfly_post	52	1.42
14	on_the_grass1	50	1.37

schedule if the user weren't interference to the character. grasp_event (event held) and fall_fountain_event (event dropped on a fountain) have appeared a lot next. These are events which happen when the user interferences to a character, i.e., reflex actions are had appeared a lot. Many find_ (event which finds something) have appeared continuously, and these are perceiving actions of a character. About active action, because many events were not able to be created, it didn't appear a lot. This result was the same in general about other characters.

In addition, we think, on reaction chaining, the character acted continuously and the system display more lifelike character. When record was investigated, about Lily, about 17% was the chin to the reaction from reaction. And, a lot of chains comparatively are the chains related to grasp.

4.3.2 Evaluation by Questionnaire

The questionnaire was also able to be taken from 20 persons who experienced the application. (See Fig. 8.) Most persons think that the operation is easy and many persons think that the operation itself is fun. Therefore it is thought that there was no problem about operation. When seeing it without interaction, the number of the

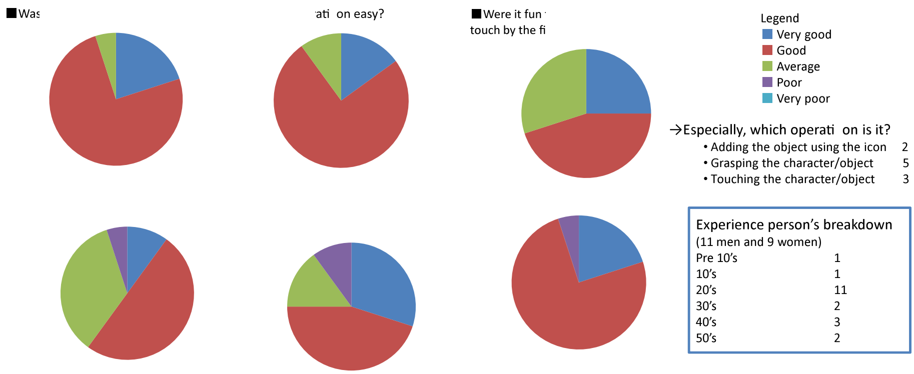


Fig. 8. Questionnaire result

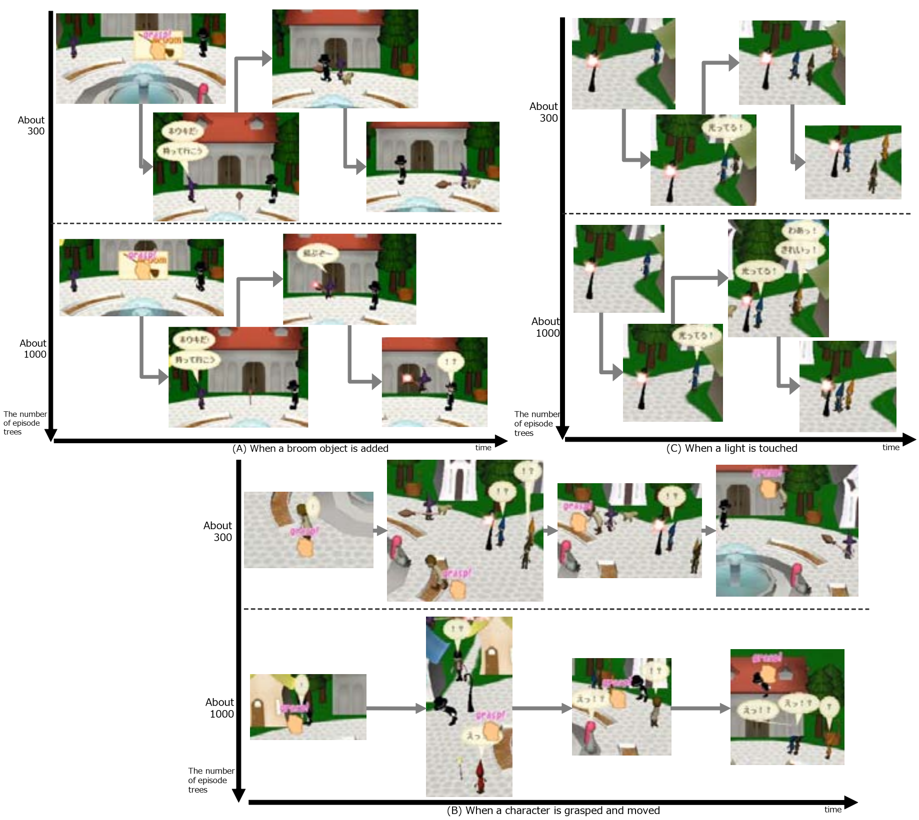


Fig. 9. Comparison of the number of episode trees, and the reactions of the characters

persons who were able to feel the character lifelike was half. If the character finishes reacting to some extent, new reaction will not be generated any more. Therefore, a possibility that the character is only taking a walk is high. The user was seeing it and replies that a character would not be lifelike. When operating, it is increasing because the character certainly react to the user interference.

5 Conclusion

In this paper, we proposed MACS, which can show characters autonomously carrying out various reactions to interference from the user in an interactive entertainment application. MACS expresses and accumulates various actions of a character with similar structures (episode trees). Thus, MACS can perform massive actions by controlling them systematically. Our experiment and exhibition demonstrated the results.

There is two directivity of scalability. The first is to make the episode tree dynamic, because a dynamic episode tree can create new storylines. The other purpose is to prepare tools to edit the episode trees visually. Because the trees are scripted by XML, we can edit them easily. However, we believe that editing will be easier if we prepare a technical graphical user interfaces.

References

1. Blumberg, M.B.: Exploring Artificial Intelligence in the new Millennium. In: Lakemeyer, G., Nobel, B. (eds.) D Learning: What Dog Learning tells us about Building Characters that can Learn. Morgan Kaufmann Publishers, San Francisco (2002)
2. Buss, A.H.: Aggression pays. In: Singer, I.J.L. (ed.) The control of aggression and violence. Academic Press, London (1971)
3. Cassell, J., Vilhjálmsson, H., Bickmore, T.: BEAT: the Behavior Expression Animation Toolkit. In: Proceedings of SIGGRAPH 2001, pp. 477–486 (2001)
4. Cavazza, M., Charles, F., Mead, S.J.: AI-based Animation for Interactive Storytelling. In: Proceedings of Computer Animation, pp. 113–120. IEEE Computer Society Press, Seoul (2001)
5. Ekman, P., Friesen, W.V.: Unmasking the Face. Prentice-Hall, Englewood Cliffs (1975)
6. Emmons, R.A.: The personal striving approach to personality. In: Pervin, L.A. (ed.) Goal concept in personality and social psychology, pp. 87–126. Lawrence Erlbaum, Mahwah (1989)
7. Funge, J., Tu, X., Terzopoulos, D.: Cognitive modeling: knowledge, reasoning and planning for intelligent characters. In: Proceedings of SIGGRAPH 1999, pp. 29–38 (1999)
8. Kopp, S., Wachsmuth, I.: Synthesizing Multimodal Utterances for Conversational Agents. The Journal Computer Animation and Virtual Worlds 15(1), 39–52 (2004)
9. Maes, P., Darrell, T., Blumberg, B., Pentland, A.: The ALIVE system: full-body interaction with autonomous agents. Computer Animation, 11–18 (1995)
10. Masuko, S., Hoshino, J.: Head-eye animation corresponding to a conversation for CG characters. Computer Graphics Forum: Journal of the European Association for Computer Graphics 26(3), 303–311 (2007)
11. Mateas, M.: An Oz-centric review of interactive drama and believable agents. Technical Report CMU-CS-97-156, School of Computer Science. Carnegie Mellon University, Pittsburgh (1997)

12. Mateas, M., Stern, A.: Façade: An Experiment in Building a Fully-Realized Interactive Drama. In: Game Developer's Conference: Game Design Track, San Jose, California (2003)
13. Meehan, J.: The metanovel: Writing stories by computer. Ph.D. Dissertation. Yale University (1976)
14. Mori, H., Hoshino, J.: Key Action Technique for Digital Storytelling. In: IFIP 4th International Conference on Entertainment Computing, pp. 36–47 (2005)
15. Murray, E.J.: Motivation and emotion. Prentice Hall, Englewood Cliffs (1964)
16. Nakano, A., Koumura, J., Miura, E., Hoshino, J.: Spilant World: Interactive Emergent Story Game using Episode Tree. The Journal of the Society for Art and Sciece 6(3), 145–153 (2007)
17. Perlin, K., Goldberg, A.: Improv: A System for Scripting Interactive Actors in Virtual Worlds. In: Proceedings of SIGGRAPH 1996, pp. 205–216 (1996)
18. Stone, M., Decarlo, D., Oh, I., Rodriguez, C., Stere, A., Lees, A., Bregler, C.: Speaking with hands: creating animated conversational characters from recordings of human performance. ACM Transactions on Graphics 23(3), 506–513 (2004)
19. Tomlinson, B., Downie, M., Berlin, M., Gray, J., Lyons, D., Cochran, J., Blumberg, B.: Leashing the AlphaWolves: mixing user direction with autonomous emotion in a pack of semi-autonomous virtual character. In: Proceedings of SIGGRAPH 2002, pp. 7–14 (2002)
20. Travers, M.: Recursive Interfaces for Reactive Objects. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 379–385 (1994)

Virtual Noctiluca: Interaction between Light and Water Using Real-Time Fluid Simulation and 3D Motion Measurement

Kyouhei Aida and Noriko Nagata

Kwansei Gakuin University, School of Science and Technology 2-1 Gakuen, Sanda, 669-1337 Japan
{aida,nagata}@kwansei.ac.jp
http://ist.ksc.kwansei.ac.jp/~nagata/

Abstract. In recent years, with the rapid improvement of the performance of computers, new possibilities for real-time simulation technologies are emerging. In this study, we simulated the behavior of water in real time and combined this simulation with a measurement of the user's 3D motion to simulate the interaction between water and light, as observed in Noctiluca. Noctiluca is oceanic plankton that produces light when physically stimulated. Stirring the surface of water containing Noctiluca in a completely dark place causes the surface to glow, and an observer can gain the mystical and fantastic experience of watching the glow becoming dim with the flow of the water.

Keywords: Interactive art, GPU, stereo vision, Smoothed Particle Hydrodynamics, animation, illumination.

1 Introduction

Simulations based on physical interactions have been widely used in computer graphics to produce animations. In particular, many methods have been developed to visualize fluids realistically simulations; however, the computational cost of these methods have always been high [1][2]. With the recent developments of techniques for obtaining faster simulations using high performance-computers, it has now become possible to calculate fluid simulations in real time. Therefore, there is a possibility that interactive applications using real-time fluid simulations will be developed soon.

In this study, we simulated the behavior of water in real time and combined this simulation with a measurement of a user's 3D motion; in this manner, we simulated the interaction between water and light, as observed in Noctiluca. Noctiluca is oceanic plankton that produces light when physically stimulated. Stirring the surface of water containing Noctiluca in a completely dark place causes the surface to glow, and an observer can gain a fantastic and almost mystical experience of watching the glow becoming dim with the flow of the water. We use Smoothed Particle Hydrodynamics (SPH), which is a method of

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 157–166, 2009.

particle-based fluid simulation, to simulate realistic water flow. Furthermore, we increased the simulation speed using graphics processing units (GPUs) so that we can receive the response quickly. We also used a stereo camera to measure the 3D motions of inputs such as a user's hand. By using 3D motions as the input data, we can experience "virtual" Noctiluca.

2 Related Work

Fluid simulation can be categorized into grid methods and particle methods. In grid methods, many grids are needed for exact computations, so it is not suited for real-time applications. In particle methods, Moving Particle Semi-implicit (MPS) [7] and Smoothed Particle Hydrodynamics (SPH) [8] are typical methods. SPH methods are suitable for real-time simulations because of their low computational cost [6].

For real-time simulations, the growth in computational power of GPUs has materially contributed to their efficiency. GPUs were originally designed for 3D graphics tasks and have also been to speed non-graphic tasks such as cellular automata simulation, particle simulation, solving of linear equations, and so on. Regarding fluid simulation, several studies have been done on increasing speed [9, 10], and Harada et al. recently proposed a method for SPH on GPUs [11].

On the other hand, several studies have been conducted on the use of physically based simulations in interactive applications. Cassinelli et al. developed an application in which a user can send portions of a projected image forward or backward in time by actually touching and deforming the projection screen, using a CCD camera [12]. Ariga et al. presented a system to show wind flow, computed by fluid simulation, projected on the screen, and then obstructed by the shadows of observers. Matsuo developed an application that controls a group of brilliant butterflies by operating a glowing ball [13]. No previous studies, however, have examined the same interactions between water and light as we propose here. In this study, integration of real-time fluid simulation and stereo vision has been realized.

3 Simulation of Water in Smoothed Particle Hydrodynamics (SPH)

3.1 Governing Equations

The velocity of a fluid is influenced by the forces associated with it, such as pressure, viscosity, and external forces. In incompressible fluids, the density remains constant even with the application of pressure. Viscosity is a force that makes the velocity constant. And the only external force is gravitational force. This indicates that the changes in the velocity of incompressible fluids is dependent on the sum of the forces associated with the fluid . The governing equations for incompressible flow are expressed by the mass conservation equation and the momentum conservation equation as follows

$$\frac{D\rho}{Dt} = 0 \tag{1}$$

$$\rho \frac{Dv}{Dt} = - \bigtriangledown p + \mu \bigtriangledown^2 v + f \tag{2}$$

where ρ, v, p, μ, and f are density, velocity, pressure, dynamic viscosity coefficient of the fluid, and external force, respectively. The left hand side of Lagrange's differential equation, indicates the changes in the physical values of the flow. It is expressed by the following equation

$$\frac{D\phi}{Dt} = \frac{\partial \phi}{\partial t} + v \bigtriangledown \phi \tag{3}$$

The second term on the right hand side of Eq.(2) indicates advection phase. When the fixed grid method is used, we must calculate the parameters associated with the advection phase. This method requires computations by linear interpolation because the grid points have physical values. This leads to serious problems in that the velocity of the fluid decreases to zero and the mass on the surface of the fluid decreases. Simulation with small time steps minimizes the problems, but this is not sufficient these simulations would fail with real-time applications. Lagrange's differential equation indicates the differential calculus on the moving particles. A particle method such as SPH, that problem does not occur because we can calculate the advection phase only by movement of particles.

3.2 Discretization

SPH is a particle method in which the fluid is regarded as a group of particle. In SPH, a physical value at a position x is calculated as a weighted sum of the physical values ϕ_j of neighboring particles j as follows:

$$\phi(x) = \sum_j m_j \frac{\phi_j}{\rho_j}(x - x_j) \tag{4}$$

where m_j, ρ_j, and x_j represent the mass, density, and position of particle j, respectively, and W is a weight function. The density of the fluid is calculated using Eq.(4) as

$$\rho(x) = \sum_j m_j \bigtriangledown W_{den}(x - x_j) \tag{5}$$

Since W is distant from the position x, the value of W becomes approximately 0. Therefore, W is the sum of the physical values of neighboring particles. In this section, we explain a method to discretize Eq.(2) with Eq.(4). In order to calculate acceleration, we separately compute pressure, viscosity, and external forces. The acceleration is then calculated as a sum of these forces. The pressure on the fluid is calculated using the constitutive equation

$$p = p_0 + k(\rho - \rho_0) \tag{6}$$

where, p_0 and ρ_0 are the pressure at rest and density, respectively. To compute the momentum conservation equation, gradient and laplacian operators, which are used to solve for the pressure and viscosity forces on particles, have to be modeled. The pressure force f^{press} and the viscosity force f^{vis} are computed as follows:

$$f_i^{press} = -\sum_j m_j \frac{p_i + p_j}{2\rho_j} \bigtriangledown W_{press}(x_i - x_j) \tag{7}$$

$$f_i^{vis} = \mu \sum_j m_j \frac{v_i + v_j}{2\rho_j} \bigtriangledown W_{vis}(x_i - x_j) \tag{8}$$

where x_i and x_j are the positions of particles i and j, respectively. The weight functions used by MNuller et al. are also used in this study [6]. The weight functions for the pressure, viscosity, and other terms are designed as follows.

$$\nabla W_{press}(r) = -\frac{45}{\pi r_e^6}(r_e - |r|)^2 \frac{r}{|r|} \tag{9}$$

$$\nabla W_{press}(r) = \frac{45}{\pi r_e^6}(r_e - |r|) \tag{10}$$

$$\nabla W_{press}(r) = \frac{315}{64\pi r_e^9}(r_e^2 - |r|^2)^3 \tag{11}$$

In these functions, the value is 0 outside the effective radius r_e.

3.3 Neighbor Search

In SPH, each particle must be used to search for neighboring particles in order to calculate the interaction among the particles. The computational cost of searching for neighboring particles is high when a large number of particles are used. To reduce this computational cost, a 3D grid covering the computational region, called a bucket, is introduced, as described by Mishra et al. [14]. Each voxel encoded as a pixel in the texture is assigned a 3D computational space. Then, for each particle, we compute a voxel to which the particle belongs and store the particle index in the voxel. If this bucket can be obtained, we do not have to search for the neighboring particles of a particle i because the neighboring particles are present in the voxels surrounding the voxel to which the particle i belongs.

3.4 SPH on Graphics Processing Units (GPUs)

To compute SPH on GPUs, physical values are stored as textures in video memories. The textures of position, velocity, density, and bucket are prepared. A

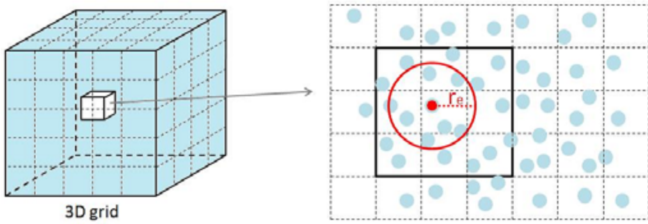


Fig. 1. Neighbor search using a bucket

texel of a texture corresponds to a particle, and the physical values of a particle are stored in the texel, as shown in Fig. 2 (a). Although a bucket is a 2D array, the current GPUs cannot write to a 3D buffer directly. Therefore, we employed a flat 3D texture in which a 3D array is divided into set of 2D arrays. We divided the 3D textures into 2D arrays, as shown in Fig. 2 (b). The indices of the neighboring particles of a particle i can be found using the generated bucket texture. Using the index of the particle, the position can be obtained from the position texture. The density of a particle i is then calculated by the weighted sum of mass of the neighboring particles, which is then written into the density texture. The pressure can be calculated from the density and position textures. Further, the viscosity force can be calculated from the position, velocity, and density textures. These forces are computed using Eq.(7) and (8). Assuming that the external force is only gravitation, we can update the velocity, and by using this updated velocity texture, the position is calculated using an explicit Euler integration as follows:

$$v_i^{t+\triangle t} = v_i^t \frac{\triangle t}{\rho_i}(f_i^{press} + f_i^{vis} + f_i^g) \tag{12}$$

$$x_i^{t+\triangle t} = x_i^t + v_i^{t+\triangle t} \triangle t \tag{13}$$

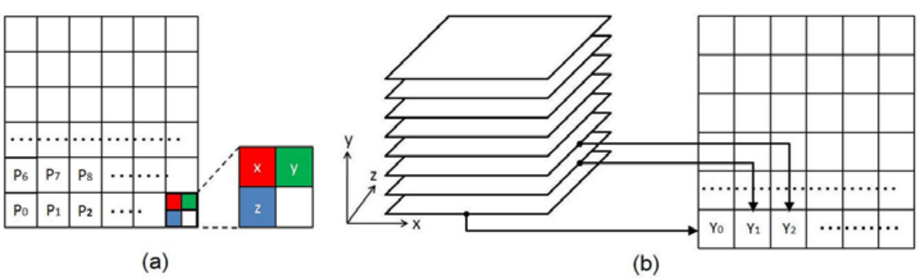


Fig. 2. (a) All the particles positions are stored at each texel of the position texture. (b) The 3D array is divided into a set of 2D arrays.

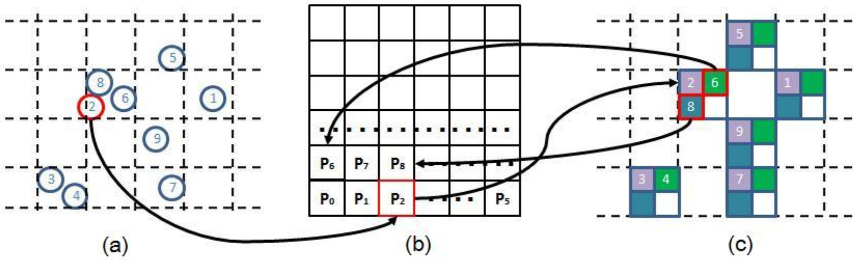


Fig. 3. Neighbor search using bucket texture. (a)disposition of particles in the region. (b) position texture (c) bucket texture comprising stored particles.

where x_i^t and v_i^t are the previous position and velocity of particle i, respectively. Although there are higher order schemes, they were not introduced because we did not encounter any stability problems.

4 Manipulation of a Stream of Water

We used a stereo camera (Bumblebee2, Point Grey Research Inc.) to measure the 3D motion of the user's hands and the parts moved by the user in real time; the 3D coordinates of the point sequence of the object surface are entered into the system. Table 1 shows the specifications of the stereo camera. The stereo camera measures the 3D points by employing the principle of stereo vision. The stereo camera cannot adjust the angle of view because the position of the two lens is fixed. However, the stereo camera can measure without calibration since it has defined parameters of the camera.

4.1 Collision Detection

In this section, we explain our method for detecting collisions between water particles, which are simulated by using SPH, and the 3D coordinates of the point sequence on the object surface. The computational cost of calculating the collision detection among particles is high, and hence, we compute it on GPUs as well as SPH. As described in chapter 3, four types of textures are prepared in SPH

Table 1. Specifications of the stereo camera

Imaging Sensor	Sony 1/3h progressive scan CCD
	ICX204 (1032 × 776 max pixels)
	4.65m square pixels
Baseline	12 cm
A/D Converter	12-bit analog-to-digital converter
Frame Rates	20 FPS
Camera Specification	IIDC 1394-based Digital Camera Specification v1.31

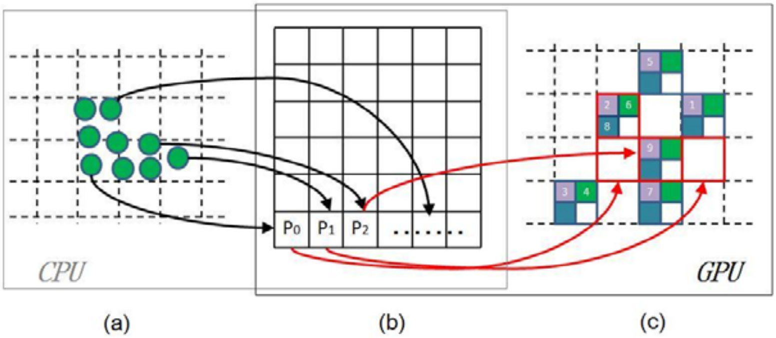


Fig. 4. Collision detection on GPUs. (a) 3D coordinates of the point sequence on the object surface. (b) Object texture. (c) Bucket texture.

on GPUs. Further, the texture of the object's position, called object texture, is prepared to compute the collision detection. The positions of the points measured by using the stereo camera are stored in each texel, and the empty texels have a value of 0. By using the object texture, we calculate the positions of the object's points and mapped them into a bucket texture. The collision detection between particles is generally computed in cases where the distance between the particles is 0 or below an optional constant value. Our method of collision detection is decided from each voxel without computing distance. We regard the water particles that are stored in the texel of the mapping object's particles as collided particles. In such a manner, we reduce the computational cost, and the method retains precision because each voxel is sufficiently small.

4.2 Taking into Consideration User's Motion

We also need to consider interference from users on the behavior of water; hence the average of all object positions is calculated for each frame. This is expressed by the following equation

$$\triangle v^t = \frac{1}{N}\sum_i x_i^t - \frac{1}{M}\sum_j x_j^{t-1}, (i = 0, \cdots, N; j = 0, \cdots, M) \tag{14}$$

where v^t is velocity of the water particle, N and M are the number of points measured by stereo camera, and x is the position of the points. The rate of change in the value is considered to be the speed of the collided water particles.

5 Result

The system was implemented using a system equipped with a Core(TM)2 CPU and a GeForce 7950GTX. The programs were written in C++ and DirectX, and the shader programs were written in HLSL. The system comprise a PC, which

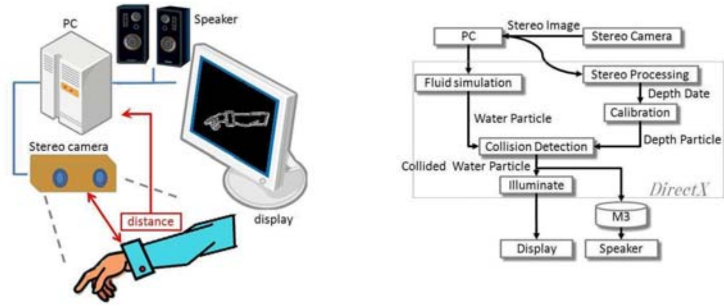


Fig. 5. Overview of the system

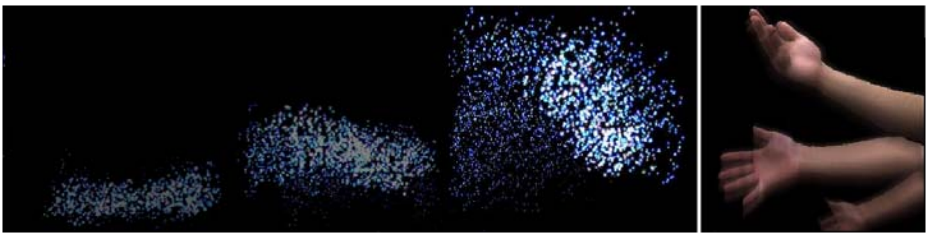


Fig. 6. The motions of a user's hand and the change in the illuminations

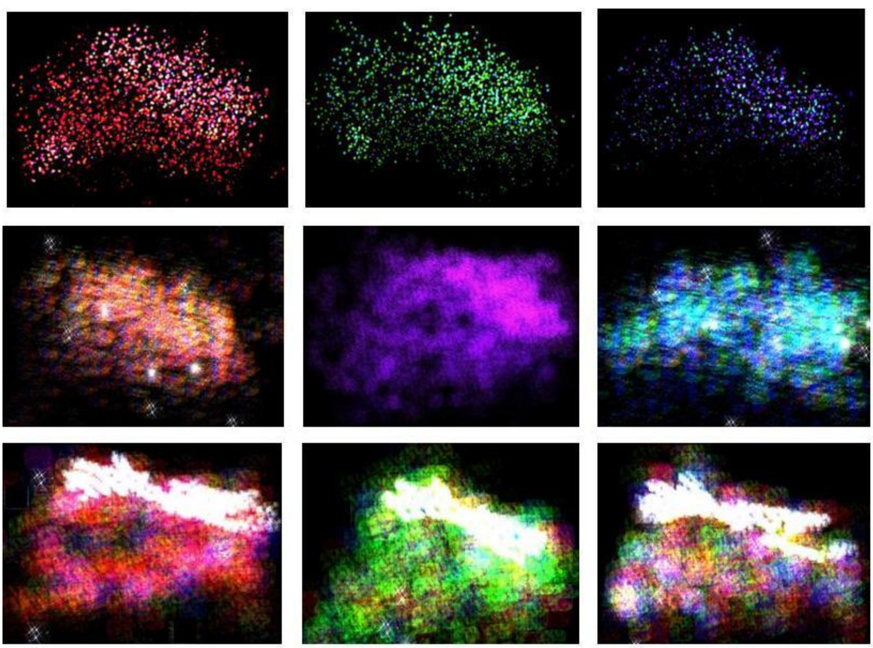


Fig. 7. Various examples of illuminations

simulates the behavior of water and regulates the overall system, and a stereo camera, which measures the user's 3D motion. Fig. 5 shows the composition of the system. Collisions between the surface point sequence and water particles were detected by the camera, and the water particles that were involved in the collision were made to produce light. In addition, in order to simplify the regulation of parameters for light production, various patterns of light production were developed by using intuitive expressions such as "beautiful", "mystical", and "fantastic". Fig. 6 illustrates a demonstration. Our system was used represent a stream of water by fluid simulation. Furthermore, we adopted sound effects that change according to the number of water particles as well as their speed, using a music visualization system [15]. This system produced synergistic effects with visual sensations and engaged users. Even with such a complex pattern of light production, we realized a frame rate of 10 fps. The system allows users to experience the natural luminescence of Noctiluca.

6 Conclusion

In this study, we have simulated the interactions of Noctiluca with water by using real-time flow simulations and 3D motion measurement. Although computational cost of fluid simulation is high, we can receive the response more quickly using GPUs effectively. Furthermore we can operate our application intuitionally measuring user's 3D motion with stereo camera. We developed the interesting application, which is not only the sense of sight, but also hearing, by adopting sound effects that change according to the user's input. Using this application, mysterious and fantastic sensations that cannot be experienced in daily life can be felt.

References

1. Irving, G., Guendelman, E., Losasso, F., Fedkiw, R.: Efficient Simulation of Large Bodies of Water by Coupling Two and Three Dimensional Techniques. In: SIGGRAPH 2006, ACM TOG 25, pp. 805–811 (2006)
2. Losasso, Shinar, T., Selle, A., Fedkiw, R.: Multiple Interacting Liquids. In: SIGGRAPH 2006, ACM TOG 25, pp. 812–819 (2006)
3. Foster, N., Metaxas, D.: Modeling the motion of a hot, turbulent gas. In: SIGGRAPH 1997: Proceedings of the 24th annual conference on Computer graphics and interactive techniques, pp. 181–188. ACM Press/Addison- Wesley Publishing Co. (1997)
4. Stam, J.: Stable fluids. In: Proceedings of the 26th annual conference on Computer graphics and interactive techniques, pp. 121–128. ACM Press/Addison-Wesley Publishing Co. (1999)
5. Carlson, M., Mucha, P.J., Turk, G.: Rigid fluid: animating the interplay between rigid bodies and fluid. ACM Trans. Graph. 23(3), 377–384 (2004)
6. Müller, M., Charypar, D., Gross, M.: Particle-based fluid simulation for interactive applications. In: Proc. of Siggraph Symposium on Computer Animation, pp. 154–159 (2003)

7. Koshizuka, S., Oka, Y.: Moving-particle semi-implicit method for fragmentation of incompressible fluid. Nucl. Sci. Eng. 123, 421–434 (1996)
8. Monaghan, J.J.: Smoothed particle hydrodynamics. Annu. Rev. Astrophys. 30, 543–574 (1992)
9. Kipfer, P., Segal, M., Westermann, R.: Uberflow: A gpu-based particle engine. In: Proceedings of the ACM SIGGRAPH/EUROGRAPHICS 6 Conference on Graphics Hardware, pp. 115–122 (2004)
10. Kolb, A., Latta, L., Rezk-Salama, C.: Hardware based simulation and collision detection for large particle systems. In: Proceedings of the ACM SIGGRAPH/EUROGRAPHICS Conference on Graphics Hardware, pp. 123–131 (2004)
11. Harada, T., Koshizuka, S., Kawaguchi, Y.: Smoothed Particle Hydrodynamics on GPUs. Computer Graphics International (2007)
12. Cassinelli, A., Ishikawa, M.: Khronos projector. ACM SIGGRAPH 2005 Emerging technologies, Article No. 10 (2005)
13. Matsuo, T.: Phantasm. In: ACM SIGGRAPH 2008 art gallery, pp. 101–101 (2008)
14. Mishra, B.K.: A review of computer simulation of tumbling mills by the discrete element method: Parti-contact mechanics. International Journal of Mineral Processing 71(1), 73–93 (2003)
15. Fujisawa, T.X., Tani, M., Nagata, N., Katayose, H.: Music Mood Visualization based on Quantitative Model of Chord Perception. J. Information Processing Society of Japan 50(3), 1133–1138 (2009)

Sound and Movement Visualization in the AR-Jazz Scenario

Cristina Portalés and Carlos D. Perales

Universidad Politécnica de Valencia, Camino de Vera, s/n. 46022 Valencia, Spain
criporri@upvnet.upv.es, carpece@doctor.upv.es

Abstract. This paper describes AR-Jazz, an augmented reality application designed to visualize sound and movements in live jazz performances. The augmented scenario is achieved within the program Max MSP Jitter, with an integrated inertial sensor and a microphone. As a display, a cinema screen is used. The application was first shown at the SedaJazz Festival 2007 in Valencia (Spain). In first place, a workshop was made in order the musicians to become familiar with the application. Secondly, a live performance was shown. The experience is described in this paper.

Keywords: augmented reality, jam session, sound visualization, inertial sensor.

1 Introduction

Augmented Reality (AR) is an emerging technology that has a great potential in several fields of knowledge, such as art [1], entertainment [2], education [3], psychology [4] and architecture [5]. This technology progresses towards collaborative and highly interactive systems, with the integration of hybrid devices that perform the 3D real-time registration. Multimodal systems are increasing, with the consideration of visual, sonic and/or haptic stimuli [6], thus enlarging the concept of AR towards more than mere visual systems.

In this paper the AR-Jazz application is described, an interactive AR system designed for live jazz performances. This paper explores the sound and movement as input data to manage the projected representation of a geometrical figure, with a group of musicians who jointly control the display of the augmented object.

This work was performed for a specific event, the Jazz & Arts Meeting inside the Panorama Jazz Festival 2007, organized by the Big Band Sedajazz and held in the auditorium of Torrent (Valencia). Jazz & Arts Meeting is an event for the interaction between Jazz and other arts such as cinema, theatre, painting, sculpture, poetry, electronics, media arts, etc., through different formative experiences and to develop creative synergies. The event is organized through a series of lectures and practices, promoting the participation of musicians and artists from other disciplines. It aims to bridge the gap between artists from different fields, to further the spread of jazz as a universal language. For this event, we prepared a workshop that was held in the morning and afternoon, and make a live performance in the evening with the students attending the workshop.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 167–172, 2009.

2 AR-Jazz in Detail

The parameters of height, intensity, timbre and spatialization are aspects that determine the sonic language. Thus, these four factors can serve as a starting point in investigating sonic augmented reality. AR-Jazz develops some of these parameters augmented in a visual projection. Physical contact of musicians with their own instruments introduces another parameter, the synaesthesia. This gesture perception in the interpretive technique is transmitted to the shape and characteristics of the projected object, creating an interaction between the musician and the computer through the sound, which may affect the interpretation of the first.

2.1 Physical Components

The elements that integrate the AR-Jazz augmented environment are (Fig. 1): A standard personal computer; The Max/MSP (v. 4.5.5) software within Jitter (v. 1.5.1); The MT9 inertial sensor of Xsens; A web cam (we used the *Live Cam! Voice* camera of Creative, within a FOV of 85º and a video resolution of 640x480 pixels); A cinema screen that acts as a mirror. The *Auditori de Torrent* screen size is of about 15x11 m; A powerful multimedia projector; A microphone; Several musical instruments.

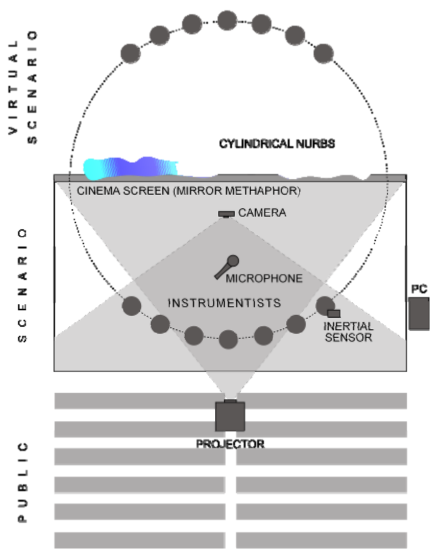


Fig. 1. Spatial arrangement of elements in the AR-Jazz application

The camera remains fixed during the performance and pointing to the musicians, so that the visual display (the cinema screen) follows the metaphor of a magic mirror (see section 2.3). The microphone is located between the camera and musicians. The spatial distribution of musicians on stage is such that, together with their own projected image, form a circle in the middle of which is the visual representation of live sound (a kind of cylindrical NURBS). The musicians give the back to the public, in such a way that they do not loose visual contact with the in real time generated virtual object. The public can see the face of the musicians from the projected image on the screen (as it were a mirror).

2.2 Virtual Elements

The produced sound is visually represented by a virtual element that consists of a modifiable cylindrical NURBS (Non Uniform Rational B-Splines), made from the Jitter object *js jitaudio2nurbs.js.* Its shape, colour, brightness and texturing varies according several characteristics of sound. Thus, the virtual element will vary depending on the timbre of each instrument, the emitted notes (height) and intensity (amplitude). Moreover, other parameters that vary randomly or at a temporary rate are introduced, as well as a change in the object orientation according to the registered rotations of the inertial sensor. The 3D position and one rotation of the object are fixed; only vary two of its rotations, which are recorded from the inertial sensor that can be carried by one of the instruments, by a dancer or by another user.

2.3 Magic Mirror Metaphor

In AR-Jazz the 'magic mirror metaphor' is introduced, as opposed to the 'magic lens metaphor', widely used in augmented reality systems. Based on the distinction between the lens and mirror, we can say that depending on the type of display used and the spatial arrangement of the camera, the augmented reality can be viewed through a window into a fictional world or as a reflection of that. The magic lens metaphor is most evident in the handheld devices, such as PDAs or mobile phones with integrated cameras, which is approximated to a lens through which we discover a new reality (the augmented environment). In AR-Jazz, the new reality is shown as an altered reflection of the real environment, reversing the spatial arrangement of the camera that shows the user as a fundamental part of the augmented environment, also increasing the sense of presence [7, 8].

2.4 Computational Processes

Jazz-AR was designed within the program Max/MSP Jitter. The human-computer interaction takes place at two levels: sonic and haptic. In the first case, the interface is a microphone that captures the sounds emitted by the musicians; in the latter case, the interface is an inertial sensor that is carried by one person (a musician, a dancer or other user). The program is composed of a principal patch and a set of sub-patches; these are described in the following lines.

The main patch is based on *jit.gl.nurbs-audiorender.pat,* which is free distributed within Jitter. The main object is *js jitaudio2nurbs.js*, which defines the shape of a cylindrical NURBS that changes its shape according to the microphone audio input. Some random values have been defined in order to change the appearence of the NURBS (blending, smooth shading, wired, surface dimensions, etc.). All the elements that define the cylinder are sent to the object *jit.gl.nurbs*.

In the *p sonido* sub-patch (Fig. 2), an analysis of the registered sound is analysed by the *fiddle~* object, obtaining the sound height and amplitude. Because these values change very quickly, the mean value of the last 20 records is considered. Then, these values are scaled and assigned to the red colour and luminosity parameters of the virtual cylinder. The values of green and blue are selected randomly.

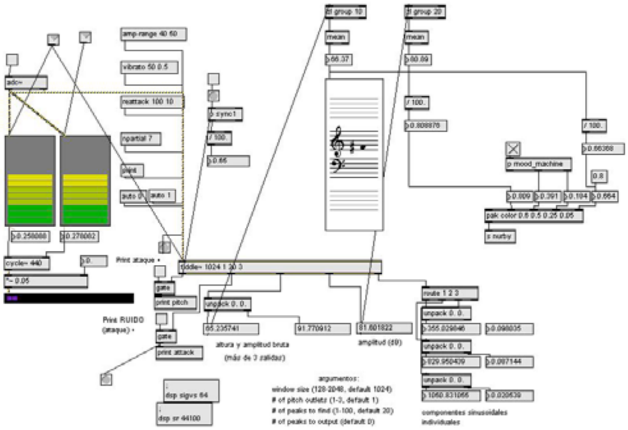


Fig. 2. Visualization of the *p sonido* sub-patch

The *p model plano* sub-patch loads a plane in *obj* format where the image captured in real time by the camera is mapped.

The *p IMU* sub-patch picks up the values registered by the inertial sensor and assigns the X and Y rotations to the virtual cylinder. This sub-patch consists of an interface between the MT9 inertial sensor and the Max/MSP Jitter software. It has been developed by Laboratorio de Luz [9] and can be freely downloaded at [10].

3 User Interaction

3.1 Workshop

The workshop was held on the stage of the auditorium for two hours in the morning and another couple of hours in the afternoon. It was attended by a total of nine musicians: a drummer, a flute, an electric guitar, a piano, an alto saxophone, two trombones, a tuba and a singer. After mounting the piece on stage, it was explained to the musicians. Then they played their instrument one by one, to see the AR-Jazz feedback to their individual interpretation. Finally several pieces of music were interpreted together in a jam session. In Fig. 3 some images are depicted.

Fig. 3. Images at the workshop

3.2 Live Performance

The live performance took place at 20:00 in the Torrent auditorium with the assistance of public. First, the *AR-Jazz* application was explained to the audience. To understand the direct relationship between the emitted sound by the instruments and the displayed graphics on the screen, each of the musicians made an improvisation with their own instrument. To make the explanation clear, at this stage the NURBS was shown on a grey background, and its rotations were fixed. In Fig. 4 some examples are depicted.

Fig. 4. Musicians playing one by one in AR-Jazz, where: a) Tuba; b) Saxo alto; c) Flute

Afterwards, some pieces were interpreted in a *Jam session* (Fig. 5). The inertial sensor was carried by one person of the audience. The represented pieces were two items that had been tested in the workshop, the first was an improvisation on the Spanish scale, whereas the second was an improvisation based on a blues rhythm.

Fig. 5. Jam session in *AR-Jazz* environment

4 Conclusions

AR can be used in many different fields; in our case we show a successful integration of this technology within the artistic discipline of jazz by means of AR-Jazz, an application designed to visualize music and movements during live performances. The use of the magic mirror metaphor allows a new distribution of musicians on the stage: although they give their back to the public, the audience is able to see their faces as a projection on a cinema screen. Interaction is achieved by means of a microphone and a miniature inertial sensor, thus integrating visual, sonic and haptic stimuli, some real and some virtually generated. Our experience shows that the use of multimodal systems does greatly enrich AR applications, increasing user and audience engagement.

Acknowledgements. This research wouldn't be possible without the collaboration of Laboratorio de Luz.

References

1. Galantay, R., Torpus, J., Engeli, M.: "living-room" Interactive, Space-Oriented Augmented Reality. In: Proceedings of the 12th annual ACM international conference on Multimedia. ACM, New York (2004)
2. Cheok, A.D., Goh, K.H., Wei, L., Farbiz, F., Fong, S.W., Teo, S.L., Li, Y., Yang, X.: Human Pacman: a mobile, wide-area entertainment system based on physical, social, and ubiquitous computing. Personal and Ubiquitous Computing 8(2), 71–81 (2004)
3. Kaufmann, H., Schmalstieg, D.: Mathematics and geometry education with collaborative augmented reality. Computers & Graphics 27(3), 339–345 (2003)
4. Juan, M.C., Alcañiz, M., Monserrat, C., Botella, C., Baños, R.M., Guerrero, B.: Using augmented reality to treat phobias. IEEE Computer Graphics and Applications 25(6), 31–37 (2005)
5. Ben-Joseph, E., Ishii, H., Underkoffler, J., Piper, B., Yeung, L.: Urban Simulation and the Luminous Planning Table: Bridging the Gap between the Digital and the Tangible. Journal of Planning Education and Research 21(2), 196–203 (2001)
6. Ha, T., Chang, Y., Woo, W.: Usability Test of Immersion for Augmented Reality Based Product Design. In: Technologies for E-Learning and Digital Entertainment, pp. 152–161. Springer, Heidelberg (2007)
7. Giner Martínez, F., Portalés Ricart, C.: The Augmented User: A Wearable Augmented Reality Interface. In: International Conference on Virtual Systems and Multimedia. Hal Thwaites, Ghent (2005)
8. Portalés Ricart, C.: Entornos de realidad aumentada en el campo del arte, in Departamento de Pintura, Universidad Politécnica de Valencia: Valencia. p. 508 (2008)
9. Laboratorio de Luz, http://www.laboluz.org/base_e.htm
10. Portalés, C.: http://personales.upv.es/criporri/multimedia_cristina_portales.php

Experimenting with Sound Immersion in an Arts and Crafts Museum

Fatima-Zahra Kaghat, Cécile Le Prado, Areti Damala, and Pierre Cubaud

CEDRIC / CNAM,
282 rue Saint-Martin, Paris, France
{fatima.azough,leprado,cubaud}@cnam.fr, areti.damala@gmail.com

Abstract. Technical museums are goods targets for experimenting with sound immersion and soundscape authoring. This paper presents an immersive sound system emitting audio content. Experimentations were conducted with a wired, proof-of-concept prototype and two wireless devices. Our system takes into consideration the position of museum visitors as well as their orientation and visual vector. In contrast with other approaches, tracking and rendering are executed locally and in real-time by the visitor's device.

Keywords: museum, immersion, edutainment, sound spatialization, head-tracking, soundscape.

1 Introduction

The project described here is justified by a simple motivation: the machines on display in the "*Musée des Arts et Métiers"* (MAM), one of the largest technical museums in France, are dumb. For many practical reasons, it is very difficult to run the machines for the public. As a consequence, visit in this type of museum turns out to be very close to a visit in a sculptures museum. The place granted to the sounds is indeed still marginal in museography and very few experiments are listed [2]. However, just like images, sounds are fundamental for learning [5]. The listening process is by nature slower compared with vision, but the reward is large since the sound is the vehicle of the human communication. The machines produce rich, complex and intense sounds often directly related with the function of the integrated mechanisms. Most machines in MAM have disappeared. However, there are strong chances that parts of the integrated mechanisms are still in use today (e.g. rods, vapor under pressure, rotating engines, etc.). The visitor could thus better comprehend the total operation of the machine exposed by associating already familiar and well known sounds. If, on the other hand, the sounds produced are not familiar, serendipity could be encouraged, with the unfamiliar acting as an element of surprise and stimulus for the visitor. Machines' sounds can be reproduced in a number of ways. In this paper, we concentrate on spatialization methods, where an auditory stimulus is positioned in virtual space defining its distance, localization (horizontal panoramic and elevation) and virtual acoustic simulation (reverberation) [3]. Real-time sound spatialization and audio augmentation is an accessible technology today, but its potential for multi-media

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 173–178, 2009.

general public edutainment applications is much less studied in comparison with visual augmentation [8].

After a review of related audio augmented environments, we summarize the useful points for our project. We then describe two experimental devices. The first one was used in an experimental setting, collecting the orientation of the head of the listener. The second is based on a commercial audio guide, in pre-production phase. We then outline the future stages of the project.

2 Related Work

Today interactive museum guides have reached a high level of functionality including visitor tracking, navigation and interaction. Bederson [1] was among the first to develop an electronic museum guide prototype supporting visitor-driven interaction by utilizing portable mini-disc players and an infra-red (IR) system to allow museum visitors to explore an exhibition at their own pace. The early European HIPS projects that run from 1998 to 2000, made also use of the IR technology. The position of the visitor was calculated through the combination of infrared and electronic compass data, then sent to a central server that pushed the appropriate information on the visitors' terminal [4]. In the LISTEN project [9] the goal was to explore immersion in audio augmented environments by overlaying a virtual soundscape to the real environment users are exploring. A tracking transmitter/receiver, based on RF-burst signals in some cases and infrared cameras for others, is integrated on a wireless headphone. A central unit collects the data of each listener such as the absolute position and the orientation, then, appropriate auditory events are selected, spatialized in real time and sent to the user headphones as a binaural data. In *ec(h)o* [6], the visitor's location is tracked using RFID technology. Sounds played are related to the objects seen by the visitor. Holding an asymmetrically shaped wooden cube, the visitor interacts with the sound objects by movement and object-based gestures, in order to listen to related audio information. Finally, Ambient Horn [5] also explored the potential of augmented audio in outdoor environments, and more in particular during a visit in woodland. The children moved to a location in which a local RF beacon was hidden; a sound was triggered and played through nearby wireless speakers while other implemented modules enabled the children to collect and exchange readings. The different architectures of ubiquitous virtual sound systems have also been discussed by Natkin et al. [7].

3 Recurrent Matters and Functional Needs

Sound information besides spoken commentaries could give to the visitors a better understanding not only of the exposed machinery but also of the MAM history. According to their movement and their behavior, the visitors receive auditory messages which can be or not related to real visual objects. There is a strong relationship between the visitor's body, the surrounding space, the time spent in a specific area and the sounds perceived. In all the previous works, the listener, the source, and the space are connected through a model of the scene, a sound map and a script which defines an interactive scenario. In order to define the script, virtual zones need to be mapped on the real space. The goal is to provide the listener with a good feeling of immersion while overlaying virtual elements inside a real scene. The coherence of an AR environment depends on the relationship between the real and virtual world and the visitor's actions. Complex auditory information is prone to hierarchical relations. This means that the content of the virtual auditory scene has to be thought as a real soundscape composition in which the listener can distinguish clearly the different components.

4 Experimentation

Our research approach is different from previous approaches in a number of ways: first in the ability to capture the visitor's head orientation, and calculate his visual vector. Then, the system is fully distributed: it provides autonomous device for each user, meaning that the management of the system is achieved locally. As a consequence, the system can be used by a large number of visitors simultaneously.

4.1 System Description

Connected to a motion and orientation sensor embedded in a headphone, the proposed system creates a map of sound objects. It constantly analyzes the visual vector of a visitor, with the aim of delivering to him the appropriate composed sound according to the objects he/she is directed to. The distance separating the visitor with each object is also taken into account. The sound intensity emitted by the exposed object and delivered to the visitor is inversely proportional to this distance. The system also includes a 3D visual interface that reproduces the museum environment, the position of the museum visitors and the sound objects around them. This interface allows the management of the museum soundscape (e.g. enabling and disabling the emitted sounds, or updating their content and nature).

Initially, our approach consisted of conceiving and carrying out a virtual simulation of a sound-guided museum exhibition. This is done by creating a mini museum environment in the lab. Each exhibited object is associated with different types of audio contents: an audio description of the object, a reproduction of the ambient sound corresponding to the object and specific musical representations. Two execution modes are proposed by the system. In the “virtual mode” the visitor's position and orientation is handled manually using the mouse and the keyboard. This mode can be used for visiting virtual museums and galleries on the net. The visitors may virtually displace themselves in the virtual environment, approach the objects and hear the audio content related to them. The “real mode” is used to visit museums in the real world. The visitor's position and orientation is handled, using a motion and orientation sensor. The visitor wears a stereo headset to which the sensor is attached (figure 2). The retrieval mechanism for the sounds object, the intensity and the orientation of each sound is based on the visitors' navigation in space which is continually updated while the visitors move through the exhibition.

Similarly, two separate graphical interfaces were attached to the system: A 3D interface and a 2D interface. The 3D interface shows the real time reproduction of the field of view of the visitor using the system. In case the virtual mode is selected, the visitor can navigate using a pointing device attached to the system (mouse, keyboard, joystick). In the 2D interface, the system displays the map of the sound objects and the position and orientation of the visitor within the museum environment. Objects are represented as visual icons while the visitor is depicted as an icon and an arrow. The interface allows then, to activate or deactivate the sound emission of a specific object by simply clicking on its associated icon. Using a "Combo Box" containing speech, ambient sound, and beeps indicators, the type of audio content can be altered.

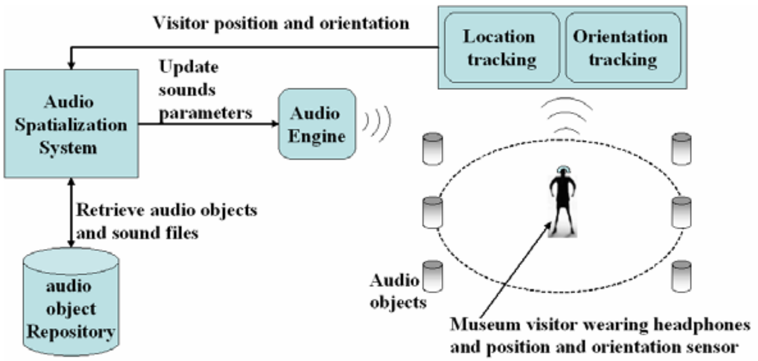


Fig. 1. System architecture

4.2 Implementation

This system is developed using the Processing environment. The FMOD API is used for the sound spatialization. It is a multi-platform sound engine, free for non commercial use. For motion and orientation tracking, a Polhemus Patriot (PP) sensor was chosen for its reliability and low latency. Using this sensor was the first step for the validation of the sound spatialization functionality. However, its wired connection is not convenient for use in large spaces. After the initial validation using the wired PP, two other wireless tracking sensors were tested: the IMU 6 Degrees of Freedom produced by Sparkfun-elecronics company and the PERCIPIO headset developed by Eshkar &Falard industrie.

Before benchmarking the system, the audio contents had to be prepared. Though the system architecture supports an unlimited number of audio objects, the issue of an optimal perception of sounds was fundamental for our system. For this reason, a test on the maximum number of distinguishable audio objects placed in the same room has been performed. The participants had to stand at a fixed point and were acoustically surrounded by a changing number of audio objects. They only had the ability to turn their heads. Most of them have been able to simultaneously perceive up to six sounds. Beyond this number, locating audio objects and distinguishing them seemed to be difficult. For this reason, the system is provided with the ability of adjusting the maximum number of audio objects to be perceived.

After choosing suitable audio samples for our experimentation, the sounds were normalized. The next step was the preparation of the experimentation environment. In a room of 4x4 meters, ten photos corresponding to audio objects were laid out on walls while the same scene was virtually reproduced in the 3D visual interface. The visitor wearing the stereo headset, on which the position and orientation sensor is

Fig. 2. (Left): User holding the PP sensor, looking towards the objects of the real scene; his field of vision is reproduced on the interface. (Right): Experimenting with the PERCIPIO headset at the CNAM museum.

fixed, is plunged into the immersive environment as soon as the audio objects are activated. His position and orientation is analyzed continuously, while the 3D characteristics of each activated sound are updated following the visitor' movements. When the visitor approaches an image, he can clearly distinguish the associated sounds according to his position and orientation (left or right). The visual field of the visitor is reproduced on the system visualization interface in real time (Figure 2). The latency between the change of user position and orientation and the update of audio content is estimated to 17 ms (milliseconds). This value is largely lower than the human auditory perception duration estimated to 50 ms (milliseconds). In addition, the visitor's head speed motion does not affect at all the auditory performance of the system.

At a second step another wireless tracking configuration was used: the industrial headset PERCIPIO. The head device is connected to a multimedia platform (PDA in our case) and can deliver personalized content according to the interests of the visitor. It is operational using the IR technology for indoor environments with 10cm precision and the GPS for outdoor environments with a 5m precision. For calculating the head orientation, PERCIPIO makes use of a magnetic compass for the determination of the azimuth angle. Presently, we are evaluating PERCIPIO in terms of feasibility, latency, accuracy while we are also working on its integration as an orientation tracker in our sound spatialization system.

5 Conclusions and Future Work

In this paper, an augmented audio reality system for experimenting sound immersion visiting a museum was presented. This system is specifically conceived and developed for technical museums, in which different types of machinery deprived of their sounds are exhibited. Our first experiments proved satisfactory in terms of performance of the tracking-rendering couple. Future work will focus on the development of a binaural individual rendering. In addition, a high acoustic quality of the adaptive sound design related to the objects of the museum, will avoid auditory strain and bring the visitor in the best perceptive conditions. Hence, the next step is to record different sounds of the machines. Then the prototype of an open composition will be created. Acquiring more data on the visitors behavior (speed, memory of trajectories, time spent in different locations) may improve this interactive scenario and the feeling of personal interactivity and immersion.

Acknowledgments

Many thanks to Bruno Jacomy, former assistant director of the Musée des Arts et Métiers in Paris, for his valuable advices, and the Eshkar company for the loan and the technical assistance regarding the Percipio headset. Thanks also to Stephane Natkin for his comments on a first draft of this paper.

References

1. Bederson, B.B.: Audio Augmented Reality: A Prototype Automated Tour Guide. In: Proceedings of Human Factors in Computing Systems, CHI 1995, pp. 210–211. ACM Press, New York (1995)
2. Jacomy, B.: L'Age du plip: Chroniques de l'innovation technique, Paris (2002)
3. Delerue, O., Warusfel, O.: Authoring of virtual sound scenes in the context of the Listen project. In: Proceedings of the AES 22nd International Conference (Virtual, Synthetic, and Entertainment Audio), pp. 39–47. AES Publications, New York (2002)
4. Oppermann, R., Specht, M.: A context-sensitive nomadic exhibition guide. In: Thomas, P., Gellersen, H.-W. (eds.) HUC 2000. LNCS, vol. 1927, p. 127. Springer, Heidelberg (2000)
5. Randell, C., Price, S., Harris, E., Fitzpatrick, G.: The Ambient Horn: Designing a novel audio-based learning experience. Personal and Ubiquitous Computing Journal 8, 177–183 (2004)
6. Wakkary, R.: Situated Play in a Tangible Interface and Adaptive Audio Museum Guide. Personal and Ubiquitous Computing Journal 11, 171–191 (2007)
7. Natkin, S., Schaeffer, F., Topol, A.: Functional Specification of a Distributed and Mobile Architecture for Virtual Sound Space Systems. In: ICMA ICMC 2001, La Havana, Cuba (September 2001)
8. Damala, A., Cubaud, P., Bationo, A., Houlier, P., Marchal, I.: Bridging the Gap between the Digital and the Physical: Design and Evaluation of a Mobile Augmented Reality Guide for the Museum Visit. In: 3rd ACM International Conference on Digital and Interactive Media in Entertainment and Arts, pp. 120–128. ACM Press, New York (2008)
9. Le Prado, C., Natkin, S.: LISTEN LISBOA: Scripting Languages for Interactive Musical Installations. In: 4th Sound and Music Computing Conference, SMC 2007, pp. 50–56. National and Kapodistrian University of Athens, Athens (2007)

BayesianBand: Jam Session System Based on Mutual Prediction by User and System

Tetsuro Kitahara[1,2], Naoyuki Totani[1],
Ryosuke Tokuami[1], and Haruhiro Katayose[1,2]

[1] School of Science and Technology, Kwansei Gakuin University
2-1 Gakuen, Sanda 669-1337, Japan
{t.kitahara,katayose}@kwansei.ac.jp
http://ist.ksc.kwansei.ac.jp/~kitahara/
[2] CrestMuse Project, CREST, JST, Japan

Abstract. One kind of pleasure that jam sessions bring is deciding a melody or an accompaniment while mutually predicting what the other participants are going to play. We propose a jam session system, called *BayesianBand*, which provides this kind of musical pleasure through sessions with computers. With this system, the chord progression in a session is not fixed in advance but rather is determined in real time by predicting the user's melody. The user, while improvising, is also expected to predict the chord progression generated by the system; accordingly, a cooperative jam session based on the mutual prediction will be achieved. To build this system, we constructed a model for melody prediction and chord inference based on a Bayesian network.

1 Introduction

The entertaining quality of music resides in the fact that music can be partly, but not fully, predicted. Predictability is indispensable for listeners' understanding of a piece of music, but if they can *completely* predict it, they cannot enjoy it. Musical pieces composed by professional musicians are therefore organized so as to achieve a satisfying tradeoff between predictability and unpredictability. Shimojo formed a hypothesis that predictability (he calls it *congruency*) in music psychologically rewards the listener for successful prediction based on his or her internal model of the music, while unpredictability (he calls it *novelty*) also brings psychological rewards, in this case as a result of the listener's detection of new information that enables his or her internal model of the music to be modified [1].

In jam sessions, these two kinds of enjoyment play an important role. During a jam session, each participant determines the melody or accompaniment to be played, while predicting what the other participants will play. When the musicians' predictions succeed and their performances sound harmonious, congruency-based pleasure (psychological reward) is obtained. When the prediction fails, but the performance nevertheless sounds harmonious, participants may attain the novelty-based psychological reward or enjoyment. The goal of

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 179–184, 2009.

our study is to provide these two kinds of enjoyment through jam sessions with computers.

In this paper, we propose a jam session system, called *BayesianBand*, in which the user and system mutually predict each other's performance. The principal feature of this system is that the chord progression, rather than being decided in advance, is decided by the system in real time. The user determines and plays the main melody by predicting what chord the system will generate in the next measure, while the system determines the next chord by predicting what melody the user will play next. Through this mutual prediction, which often succeeds but sometimes does not, the user can obtain the two kinds of musical pleasure described by Shimojo [1].

2 Technical Requirements and Related Work

BayesianBand is a jam session system that determines a chord progression in real time by predicting the user's melodies. The purpose of this system is to provide the user with both the congruency- and novelty-based enjoyment through jam sessions. To obtain the congruency-based enjoyment, the input (the user's melody and previous/current chords) and the output (the subsequent chord) should have a causality. For this reason, we do not introduce any randomness into the determination of the output from the input, even though this is a common approach to maintaining novelty[2]. In addition, this causality should be acquired by users through jam sessions and thus should be consistent and musically appropriate. If the causality is completely immobilized, however, users may discover the input/output relationships completely and quickly become bored with the jam session. The causality should therefore always evolve. To summarize, causality between input and output should be (1) deterministic (not involving any random process), (2) musically appropriate, and (3) always evolving, in order to attain both congruency and novelty.

To fulfill these three requirements, in BayesianBand, we use a probabilistic model for melody prediction and chord inference, as follows:

1. The chord having the maximum likelihood, given an input, is always determined.
2. The probabilistic model is trained with existing pieces of music.
3. The probabilistic model is incrementally updated to adapt to the user's melodic tendency.

Various jam session systems, described in previous studies [3,4], have been developed, but most of these assumed that the chord progression is fixed in advance. Aono et al. [5] developed a jam session system that did not assume a fixed chord progression. When the user plays a chord progression (the system judges so if more than three notes are played simultaneously), the system automatically recognizes it and considers it to be repeated. This system therefore did not aim to determine the chord progression by predicting the user's performance.

Melody predictions have also been widely attempted. Conklin [6] developed a melody prediction system by regarding melodies as Markov chains. Pachet [2] developed a system, called the *Continuator*, that generates a sequence of notes that can follow the melody played by the user. This system learns the user's melodies as a Markov tree structure and, when the user plays a melody, recursively generates the following note, based on this structure. There have thus been various studies of Markov-based melody prediction, but they did not model the horizontal dependency of the chord progressions behind melodies.

Harmonization, which aims to give a chord progression to a melody, is also an important topic and has been widely studied. Kawakami et al. [7] proposed a method for harmonization using a hidden Markov model (HMM) in which the melody and chord progression are modeled as observed and hidden variables, respectively. These studies, however, assume that the whole input melody is referable from beginning to end; they do not aim at harmonization for the future's melody by predicting it.

3 System Overview and Algorithm

The main functions of BayesianBand are (1) the chord determination by prediction of the user's melody and (2) incremental updating of the prediction model. In this section, we describe their algorithms used to generate these functions after providing the system overview.

3.1 Problem Statement

The input is the user's melody; the output is a chord progression. For simplicity, the timings of chord changes are limited to occurrence at the beginning of each measure. The input is a monophonic melody, the key is given, and no key modulation occurs. The first chord is the tonic chord of the given key. Due to a small amount of training data, the target chords are limited to the seven diatonic chords.

3.2 System Overview

The system overview of BayesianBand is shown in Fig. 1. Because the initiative for tempo control is governed by the system, the accompaniment is automatically performed with a constant tempo. When the user presses a key on the MIDI keyboard, the system predicts the next note and infers the next chord. This process is repeated for each keystroke; accordingly, the chord inference result is updated after each keystroke. Immediately before changing the measure, the chord having the maximum likelihood is determined as the next chord. In parallel, the incremental update of the prediction model is performed at each keystroke.

3.3 Algorithm for Melody Prediction and Chord Determination

Here we deal with the problem of inferring the most likely subsequent chord c_{t+1} of a chord progression $\boldsymbol{c} = (c_1, \cdots, c_t)$ by predicting the next note n_{t+1} of

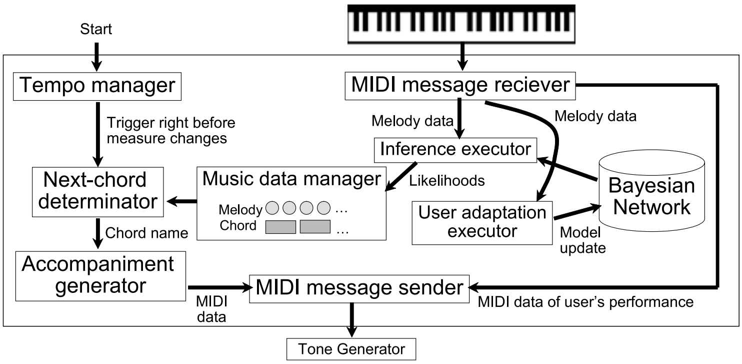
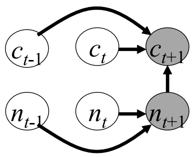


Fig. 1. System overview of BayesianBand



n_i	12 note names
c_i	7 diatonic chords

Fig. 2. Bayesian network used for melody prediction and chord determination

a given melody $\boldsymbol{n} = (n_1, \cdots, n_t)$. In general, a melody and a chord progression have different sequential causalities, $p(n_{t+1}|\boldsymbol{n})$ and $p(c_{t+1}|\boldsymbol{c})$, and simultaneous elements in the melody and chord progression also have a causality, described as $p(c_t|n_t)$. For simplicity, the sequential causalities $p(n_{t+1}|\boldsymbol{n})$ and $p(c_{t+1}|\boldsymbol{c})$ are approximated by trigram models, described as $p(n_{t+1}|n_{t-1}, n_t)$ and $p(c_{t+1}|c_{t-1}, c_t)$, respectively.

Using these sequential and simultaneous causalities, the relationship between a melody and a chord progression can be described as the Bayesian network shown in Fig. 2. In general, Bayesian networks should be singly connected because when this is the case, a low-complexity algorithm for probability calculation can be applied. To make a singly-connected network, the dependencies unrelated to the nodes n_{t+1} and c_{t+1} are omitted, since the values of the other nodes have already been observed or determined.

The inference process is performed at each keystroke. Once a key is pressed, the observed note names are set to n_{t-1} and n_t and the determined chord names to c_{t-1} and c_t. Then the inference is executed: the probability densities for the nodes n_{t+1} and c_{t+1} are calculated using Pearl's method [8]. After this process is repeated at each keystroke, the chord having the maximum likelihood is determined as the next chord immediately before changing the measure.

3.4 Algorithm for Incremental Model Update

Incremental model update aims not only to retain novelty in chord determination but also to improve the accuracy of melody prediction by adapting the model to the user. The basic method of achieving this is to calculate the conditional probability $p(n_{t+1}|n_{t-1}, n_t)$ as a weighted mean of the probability precalculated from a corpus and the probability calculated online from the user's performance. As the number of notes performed by the user increases, the weights are gradually changed so that the latter's weight becomes larger. Specifically, the conditional probability $p(n_{t+1}|n_{t-1}, n_t)$ is calculated using the following equation:

$$p(n_{t+1}|n_{t-1}, n_t) = \frac{p_0(n_{t+1}|n_{t-1}, n_t) + \alpha \{\log N(n_{t-1}, n_t)\} \dfrac{N(n_{t-1}, n_t, n_{t+1})}{N(n_{t-1}, n_t)}}{1 + \alpha \log N(n_{t-1}, n_t)},$$

where $p_0(n_{t+1}|n_{t-1}, n_t)$ is the probability calculated from a corpus, $N(n_{t-1}, n_t)$ is the frequency that the user played n_{t-1} and n_t in this order, $N(n_{t-1}, n_t, n_{t+1})$ is the frequency that the user played n_{t-1}, n_t, and n_{t+1} in this order, and α is a constant.

4 Implementation and Trial Use

4.1 Implementation

We implemented a prototype system of BayesianBand using Java. We used CrestMuseXML Toolkit[1] for implementing the overall framework and Weka[2] for learning and using the Bayesian network. For learning the Bayesian network, we used 415 pieces of standard jazz music (pairs of melodies and chord progressions).

4.2 Results of Trial Use

The first author used the implemented prototype system. After repeating a jam session several times, he understood rough trends in chord progressions generated by the system to some extent and reflected his chord prediction in his improvisation. He felt pleasure when the predicted chord was actually played by the system. When his chord prediction failed, his melody and the generated accompaniment often sounded unharmonious, but in some cases they sounded harmonious. In those instances, he felt novelty. Thus, BayesianBand to some extent successfully provided a trial user with the two kinds of enjoyment.

Predicting the next chord while playing was enjoyable, like a game in itself. This type of pleasure cannot be provided by jam sessions in which the chord progression is fixed in advance or is determined at random.

An excerpt of the melodies played and the prediction results are shown in Fig. 3. When the three best candidates for each note prediction were considered, the notes immediately following half of the played notes were successfully predicted.

[1] http://www.crestmuse.jp/cmx/

[2] http://www.cs.waikato.ac.nz/ml/weka/

Melody prediction			Chord inference	
C (.23)	E (.15)	D (.15)	F (.42)	C (.16)
C (.30)	**E** (.26)	G (.18)	F (.45)	Em (.16)
D (.25)	G (.25)	C (.16)	F (.41)	Dm (.16)
C (.45)	G (.16)	B (.13)	F (.48)	Em (.16)
G (.21)	**E** (.17)	B (.13)	F (.34)	Em (.18)
C (.47)	**G** (.19)	F (.08)	F (.50)	Dm (.15)
C (.34)	B (.34)	D (.07)	F (.39)	Em (.22)
A (.28)	E (.20)	G (.17)	F (.30)	Em (.22)
F (.16)	E (.16)	A (.13)	F (.32)	Em (.18)
E (.33)	**A** (.23)	F (.13)	F (.33)	Em (.20)
D (.22)	G (.22)	C (.20)	F (.30)	C (.22)
E (.43)	**C** (.25)	G (.11)	F (.29)	C (.24)
A (.33)	G (.19)	**E** (.13)	F (.31)	Em (.21)
C (.47)	G (.19)	F (.08)	F (.29)	C (.25)
C (.47)	G (.19)	F (.08)	C (.30)	F (.19)
C (.56)	G (.10)	E (.10)	C (.32)	F (.20)
A (.27)	D (.17)	E (.15)	Am (.22)	F (.21)
D (.22)	G (.22)	C (.19)	Am (.21)	C (.20)
E (.43)	C (.25)	G (.11)	Am (1.0)	
C (.38)	Others (.06)		C (.28)	F (.19)
B (.32)	C (.14)	E (.14)	Am (.22)	C (.17)
C (.51)	G (.11)	D (.09)	Dm (.23)	Am (.14)
G (.21)	E (.17)	B (.13)	Dm (.16)	Am (.15)
C (.40)	A (.30)	**E**,G (.07)	Dm (.22)	Am (.14)
C (.36)	E (.14)	D,G (.11)	Dm (.21)	Am (.15)

Fig. 3. Example of performed melodies and its melody-prediction and chord-inference results obtained using BayesianBand. The values in parentheses are the likelihoods; the boldfaced characters represents the names of the notes that were actually played.

5 Conclusion

In this paper, we proposed a new jam session system, called *BayesianBand*, in which the user and the system mutually predict each other's performance. The topic of human-system jam sessions based on the mutual prediction is interesting as a target domain for research involving man-machine collaboration. In the future, we plan to investigate through long-term experiments how humans and systems collaboratively create music.

References

1. Shimojo, S.: Research plan for Shimojo implicit brain function project, `http://impbrain.shimojo.jst.go.jp/jpn/about_jpn.html` (in Japanese)
2. Pachet, F.: The Continuator: Musical interaction with style. In: Proc. ICMC (2002)
3. Nishijima, M., Watanabe, K.: Interactive music composer based on neural networks. In: Proc. ICMC, pp. 53–56 (1992)
4. Goto, M., Hidaka, I., Matsumoto, H., Kuroda, Y., Muraoka, Y.: A jam session system for interplay among all players. In: Proc. ICMC, pp. 346–349 (1996)
5. Aono, Y., Katayose, H., Inokuchi, S.: An improvisational accompaniment system observing performer's musical gesture. In: Proc. ICMC, pp. 106–107 (1995)
6. Conklin, D., Witten, I.H.: Multiple viewpoint systems for music prediction. J. New Music Res. 24, 51–73 (1995)
7. Kawakami, T., Nakai, M., Shimodaira, H., Sagayama, S.: Hidden markov model applied to automatic harmonization of given melodies. IPSJ SIG Notes, 99-MUS-34, 59–66 (2000) (in Japanese)
8. Pearl, J.: Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, San Francisco (1988)

v.morish'09: A Morphing-Based Singing Design Interface for Vocal Melodies

Masanori Morise[1], Masato Onishi[2], Hideki Kawahara[3], and Haruhiro Katayose[4]

[1] College of Information and Acience, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577 Japan
[2] Graduate School of Systems Engineering, Wakayama University, 930 Sakaedani, Wakayama 640-8510, Japan
[3] Faculty of Systems Engineering, Wakayama University, 930 Sakaedani, Wakayama 640-8510, Japan
[4] School of Science and Technology, Kwansei Gakuin University, 2-1 Gakuen, Sannda, Hyogo, 669-1337 Japan
http://crestmuse.jp/index-e.html

Abstract. This paper describes a singing design method based on morphing, the design and development of an intuitive interface to assist morphing-based singing design. The proposed interface has a function for real-time morphing, based on simple operation with a mouse, and an editor to control the singing features in detail. The user is able to enhance singing voices efficiently by using these two functions. In this paper, we discuss the requirement for an interface to assist in morphing-based singing design, and develope an interface to fulfill the requirement.

Keywords: Singing voice synthesis, voice morphing technique, user interface design.

1 Introduction

Vocal manipulation is one of the most entertaining elements of computer music processing. Above all, replacing the singing style or voice characteristics of a singing voice with those of a professional singer has been a highly desirable function in vocal manipulation applications.

In the field of sound synthesis study, vocal synthesis has been a major research target, and core technologies have already been proposed [1],[2]. Commercial vocal synthesizers for PCs were released once computer and media technology had developed sufficiently, by 2005. In 2007, a vocal synthesizer, the conceptual basis of which is a virtual animation vocalist, *Hatsune Miku* (Vocaloid2), was released. As of late 2008, sales of Hatsune Miku had reached 40,000, a record-breaking number in desktop music software sales. Amateur creators are uploading their original songs, synthesized using Hatsune Miku, to video-sharing websites, and some of these videos have ranked in the top 10 most viewed. Users of vocal

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 185–190, 2009.

synthesis software edit lyrics and melody, i.e., pitch trajectory, using an editor that resembles a music sequencer. Some vocal synthesizing software provides functions for automatic control of delicate pitch trajectories, such as portamento and vibrato, but users are obliged to elaborate on the pitch control parameters in order to obtain human-sounding singing voices. This operation is so troublesome that tools are being proposed for adjusting the parameters of vocal synthesizers so that the output resembles human singing voices [3].

We have been developing a morphing-based singing design technology capable of replacing the singing style or voice characteristics of a singing voice with those of a professional singer. In this paper, we discuss technologies and interfaces that allow the user to generate a newly synthesized voice by adjusting two morphing rates: one for pitch changes and one for voice characteristics.

The rest of this paper is organized as follows: Section 2 presents singing voice morphing and peripheral technology. Section 3 discusses requirements for morphing-based singing design and the development of *v.morish'09* to fulfill these requirements. Finally, Section 4 concludes this paper.

2 Singing Design

In this section, we discuss a new parameter for use in singing design, singing voice morphing [4], and a peripheral technology, called STRAIGHT [5]. Generally, to facilitate understanding when we discuss singing voices, we use an example of a well-known singer instead of talking about the fundamental frequency (F0) or spectrum envelope. Although singing voices can be analyzed in terms of their F0 and spectrum envelope, it is difficult for us to explain the allure of the singing voice by citing these properties. Instead, we talk about the singing style and voice characteristics of a well-known singer. Therefore, although most vocal synthesizing software has an editor to control F0 and the spectrum envelope as the main parameters, it is difficult for creators to synthesize a desired singing voice by controlling these parameters. Synthesizing a natural singing voice might be simpler if the user could control a singer's style and voice characteristics instead of the F0 and spectrum envelope.

We propose a technique for singing voice morphing that controls singing style and voice characteristics by morphing the user's singing voices with a singer's characteristics. This technique enables the user to blend two features (singing style and voice characteristics) independently. The morphing of two singers' voices has also been proposed for Karaoke applications [1], but singing voice morphing differs from such applications in that there are two control parameters. A high-quality vocoder, STRAIGHT [5], is used for the singing voice analysis, morphing and synthesis. The user is able to model the singing voices by controlling the morphing rates of the singing style and voice characteristics. Furthermore, singing voice morphing also enables the user to control emotional color by blending emotional singing voices into the user's own singing voices.

The interface for STRAIGHT-based singing voice morphing consists of mapping the singing style and the voice characteristics to the horizontal and the

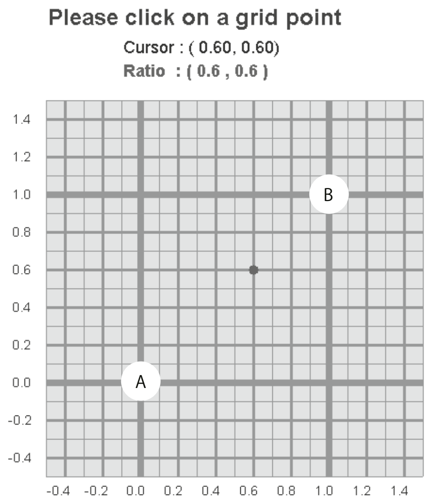


Fig. 1. GUI for singing voice morphing. The horizontal axis represents the morphing rate for the singing style. The vertical axis represents the morphing rate for the voice characteristics.

vertical axes of a two-dimensional plane (Figure 1); the user is able to identify the morphing rates visually. This interface enables the user to reproduce singing voices with morphing rates shown by the cursor. However, because STRAIGHT cannot achieve real-time analysis, morphing and synthesis, the user cannot change morphing rates during reproduction. In this paper, we use TANDEM-STRAIGHT[6] that produces the same results that STRAIGHT does, only much more quickly.

2.1 STRAIGHT and TANDEM-STRAIGHT

STRAIGHT, which is a vocoder system [7], analyzes a voice in terms of its fundamental frequency (F0), spectrum envelope and aperiodicity spectrogram (Figure 2). It is able to synthesize a voice that sounds as natural as a human voice captured by a microphone. In singing voice morphing, F0 is represented by the term "singing style." The spectrum envelope and aperiodicity are represented by "voice characteristics." General voice morphing [8] uses one control to morph these three parameters, and singing voice morphing has two parameters, corresponding to singing style and voice characteristics. STRAIGHT is incapable of real-time morphing and synthesis because they would require too much computational power. Therefore, the interface, shown in Figure 1, can only reproduce singing voices morphed in advance. The development of a real-time interface based on STRAIGHT has been difficult.

TANDEM-STRAIGHT [6] produces the same results that STRAIGHT does, only much more quickly. The problem of real-time synthesis is solved by using

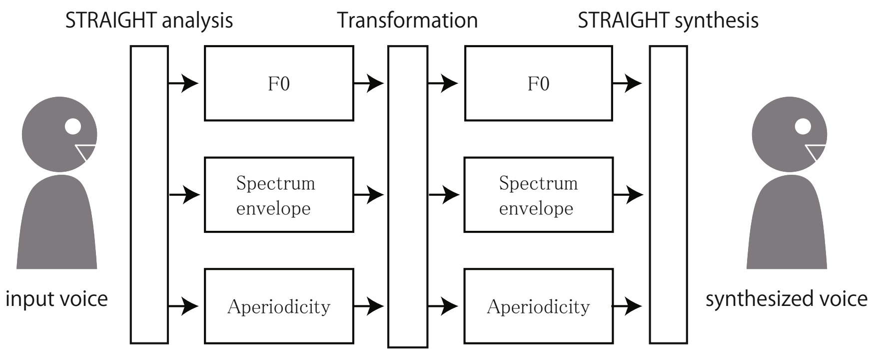


Fig. 2. Overview of STRAIGHT (TANDEM-STRAIGHT works the same way)

TANDEM-STRAIGHT. We have also developed a library in the C programming language, so that a real-time interface might be developed for TANDEM-STRAIGHT. This library enables us to change morphing rates in real time during reproduction.

3 Design and Development of the Singing Design Interface

The interface supports morphing-based singing design and provides a method for aligning two singing voices. Most vocal synthesizing software is able to show the results of the user's input. Singing voice morphing must display the results of changing morphing rates so that the user will be able to perceive the difference in the quality caused by changing the morphing rate.

It is essential that the software include an off-line editor for editing singing voices. Previous software incorporates an off-line editor to control each parameter in singing design. Therefore, our interface will have an editor for controlling the morphing rates of singing style and voice characteristics. The user will be able to create the detailed results with this editor.

3.1 Implementation of *v.*morish'09

We developed *v.morish'09* as an interface to fulfill the requirements described in section 3. Figure 3 is a screenshot of v.morish'09.

In Figure 3, the left-hand side of the interface provides real-time morphing control. The horizontal axis represents the singing style, and the vertical axis represents the voice characteristics (here, "voice color"). With this interface, the user can control morphing rates in real time during reproduction of the morphed singing voice, making it possible for the user to hear the changes being made. The right-hand side of Figure 3 shows an off-line editor for drawing the morphing rates of the singing style and the voice characteristics. The user is able to draw the morphing rate trajectory in detail using this off-line editor.

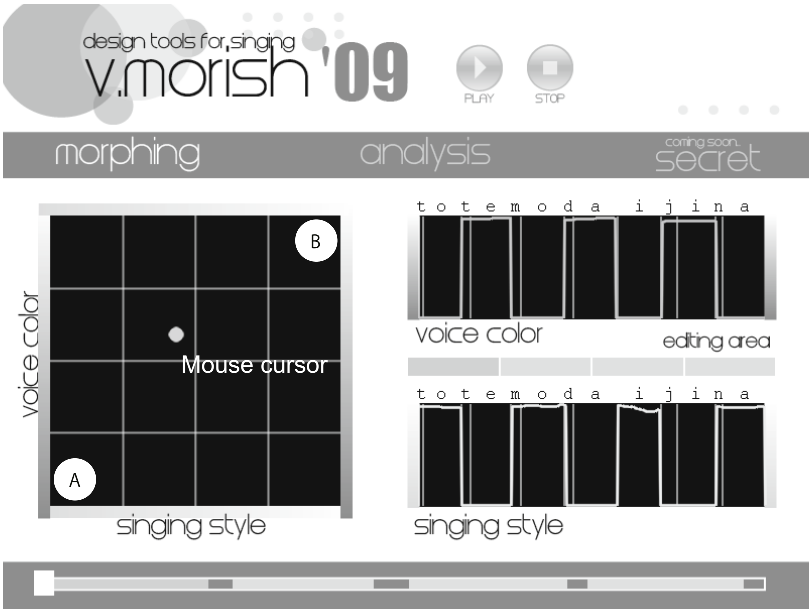


Fig. 3. Screenshot of the morphing-based singing design interface *v.morish'09*. This interface consists of two tools: a real-time GUI for morphing (left) and an editor for off-line processing (right).

Additionally, v.morish'09 can reproduce the morphed singing voices as modified by the morphing rate trajectory in the editor. Thus, the user can design singing voices by using these two functions.

3.2 Two Procedures for the Singing Voice Design

The user designs singing voices using the real-time interface both to draw a rough trajectory and to revise this trajectory.

Real-Time Control Using the Mouse. The movement of the cursor during reproduction is reflected in real time by the morphed results. The change of the morphing rate is also shown in the trajectory represented within the editor. Therefore, detailed, real-time control is possible with this editor.

Detailed Design Using the Editor. The rough trajectory drawn in real time can be revised with the editor. The editor also shows phoneme boundaries, providing control over the morphing rates of each phoneme. Moreover, v.morish'09 is able to reproduce the morphed singing voices according to the morphing rate trajectory shown in the editor. The real-time interface cannot adjust the morphing rate trajectory in detail and reproduce by the same trajectory. Using the reproduction function, the user is able to reproduce the morphed singing voices according to the same trajectory.

3.3 Discussion

Many singing design methods and singing design interfaces have been proposed in the past. The proposed interface is different from them in control parameters that enable users to change singing style and voice characteristic directly. Replacing the singing style or voice characteristics of a singing voice with those of a professional singer will be possible by using v.morish'09. As v.morish'09 has the real-time control GUI and the editor to draw trajectory, the user may control these parameters easily. Evaluation of usability is important future work.

4 Concluding Remark

In this paper, we discussed the requirements for morphing-based singing design, and proposed the interface v.morish'09 as a support tool of for morphing-based singing design. Our v.morish'09 enables users to control the singing style and voice characteristic directly. The evaluation of usability of v.morish'09 is important future work.

Acknowledgments. This research was partly supported by the CrestMuse project, conducted by the Japan Science and Technology Agency (JST) and a grant-in-aid for young scientists (Start-up) 20800062.

References

1. Cano, P., Loscos, A., Bonada, J., de Boer, M., Serra, X.: Voice morphing system for impersonating in karaoke applications. In: Proc. ICMC, pp. 109–112 (2000)
2. Bonada, J., Serra, X.: Synthesis of the singing voice by performance sampling and spectral models. IEEE Signal Processing Magazine 24, 67–79 (2007)
3. Nakano, T., Goto, M.: VocaListener: An automatic parameter estimation system for singing synthesis by mimicking user's singing. In: Proc. IPSJ SIGMUS meeting, May 2008, pp. 49–56 (2008) (in Japanese)
4. Kawahara, H., Ikoma, T., Morise, M., Takahashi, T., Toyoda, K., Katayose, H.: Proposal on a morphing-based singing design manipulation interface and its preliminary study. Journal of Information Processing Society 48, 3637–3648 (2007) (in Japanese)
5. Kawahara, H., Masuda-Katsuse, I., de Cheveigné, A.: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction. Speech Communication 27, 187–207 (1999)
6. Kawahara, H., Morise, M., Irino, T., Takahashi, T.: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In: Proc. ICASSP, pp. 3933–3936 (2008)
7. Dudley, H.: Remaking speech. J. Acoust. Soc. Am. 11, 169–177 (1939)
8. Kawahara, H., Nisimura, R., Irino, T., Morise, M., Takahasni, T., Banno, H.: Temporally variable multi-aspect auditory morphing enabling extrapolation without objective and perceptual breakdown. In: Proc. ICASSP 2009, pp. 3905–3908 (2009)

New Hitch Haiku: An Interactive Renku Poem Composition Supporting Tool Applied for Sightseeing Navigation System

Xiaofeng Wu[1], Naoko Tosa[1], and Ryohei Nakatsu[2]

[1] Academic Center for Computing and Media Studies, Kyoto University
Yoshida-Nihonmatsu, Sakyo, Kyoto, 606-8501 Japan
XiaofengWu@hotmail.com, tosa@media.kyoto-u.ac.jp
[2] Interactive & Digital Media Institute, National University of Singapore
Blk E3A #02-04, 7 Engineering Drive 1, Singapore 117574
idmdir@nus.edu.sg

Abstract. As is well-known, cultures are rooted in their unique regions, histories and languages. Communication media have been developed to circulate these cultural characteristics. As a part of our research "Cultural Computing", which means the translation of cultures using scientific methods representing essential aspects of Japanese culture[1], an interactive Renku poem generation supporting system was developed to study the reproduction of a traditional Japanese Renku by computer. This system extended the functionality of our previous Hitch-Haiku system to the Renku based on same association method and attached more cultural characteristics on it: the Renku verse displayed on the Japanese-style color pattern which represents the same season in Renku Kigo (seasonal reference) and the generated Renku verse including the information of sightseeing place.

Keywords: Haiku, Renku, Renku generation, Interactive art, Association.

1 Introduction

Haiku is a Japanese traditional poem style with minimal length of seventeen syllables (in Japanese) in three metrical phrases including a seasonal word called "Kigo." The original form of Haiku was called Hokku and in the late 19th century Shiki Masaoka revised it and finally established the present form of Haiku [3]. Haiku include various imaginative expressions and thus has been applauded by many people. Haiku is a story that generates context - the shortest story in the world. Known as the first great Haiku poet in the Japanese history, Matsuo Basho is responsible for "Oku No Hosomichi", a prime example of his work [4].

In 1959, Theo Lutz developed a system of a poem generation for the first time [5]. The system only showed words at random on grammatical rule, and could not generate a poem with user's interactions. In 1971, for the first time Masterman developed the generation system of a Haiku[6]. By rearranging the words which users chose from the pull down menu in the interaction process the system generates a Haiku. However, in these interactions, users could input only a few limited words into the system.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 191–196, 2009.

In the field of Interactive Art or Game, the quality of contents is important, like *"Passage Sets / One Pulls Pivots at the Tip of the Tongue"* by Bill Seaman [7], and *"An Anecdoted Archive from cold War"* by George Legrady[8]. But from technologies viewpoint, only simple techniques have been used. On the other hands, in the field of AI, many researchers and developers have been using various kinds of techniques to find some relations among input words/phrases by users and to compose answers in relation to these inputs [9]. These techniques have been often used, because using one of these techniques they can develop an interactive system that can achieve relatively interesting interactions. But the relations they try to find out and they try to use in their systems are static, and the quality of their interactions have been mostly dependent on the quality of the relations given beforehand.

Based on his long carrier and an editor and a philosopher, Matsuoka defined four several basic forms called "*Thoughtforms*," as basic form of relations among things [10]. As the "*Thoughtforms*" can work as the basic method to re-construct interesting relations among words and enable to generate better Haiku, Tosa applied the technique to develop an interactive system, "Hitch Haiku", which supports a user for composing a Haiku [11]. The user only need to input some words into the system, and the system can compose phrases consisting of five-seven-five syllables which most fit to the user inputs. The system is called Hitch Haiku as it generates a Haiku "hitching" the phrases chosen based on the user inputs.

In order to extend this Hitch Haiku system and apply it to wider field, we adopt the success of the previous system, that is, generation of better quality Haiku using association method, and use more cultural features to develop a new interactive Renku composition supporting system which is applied for the Kyoto sightseeing navigation system.

2 Concept

The new Hitch Haiku system is designed to be a Renku composition supporting subsystem applied in the sightseeing navigation system which is based on our previous Hitch Haiku system, an interactive Haiku generation supporting system. Compared with the old Hitch Haiku system, there are three unique features in our new Hitch Haiku system:

(1) To generate Renku verse using searching and association method, the latter is the major point to get the better quality of Renku verse

The rules of Renku is much more restricted and complicated than those of Haiku, especially the limitation usage of Kireji in Renku creation. Based on the experience from previous experiment that the quality of the Haiku generated depends on the sensitivity of the user input words, we give up the steps of attaching Kireji to user input words to generate the phrases and try to generate the phrases of Renku verse in the way like human done, i.e., create the haiku phrase by the meaning of previous generated phrase and new keyword which can correspond to the previous generated phrase.

(2) During the generation of Renku verse, optionally we applied the Haiku thesaurus words related to the sightseeing place. Thus, it is possible to include the features of the specific sightseeing place into the Renku verse.

(3) To show generated Renku verse on the Japanese-style color pattern, and furthermore, the selection of background color pattern is based on the representative season which corresponds to the Kigo of the generated Renku verse. Therefore, the user can feel the cultural characteristics not only from the phrases of Renku verse but also from the background color pattern which is also harmonious to the Renku verse.

By adding above features into our new Hitch Haiku system, we hope that the generated Renku verse and display style can show users more cultural characteristics.

3 Process of Generation

The Renku generation procedure in our new system is the process shown in follows (Fig.1). In details, the system works in the following steps:

(1) A user input his or her two favorite words (or phrases) in a text box of the Hokku (the first Haiku of the Renku) input webpage and select the sightseeing place in the Kyoto city from a optional selection box which including the most famous 110 hot spots of this city using mobile phone.
(2) The system searches the Haiku phrase databases both with and without Kigo and found out the phrases including one of the user input word and randomly select one. For Hokku, the system will do a special selection procedure: prior to select the phrase with Kireji. For the other Renku verses, because the Kirejis are forbidden, the system will select the phrase without any Kireji.
(3) In the other hand, the system searches the Haiku thesaurus word related to the user chosen sightseeing place. This Haiku thesaurus word will be used in the later Renku phrase generation.
(4) Once the first phrase was selected, a syntactic analysis for this phrase is carried out in the web server and the basic forms of noun or verb from each phrase are extracted.
(5) Then the system apply the elemental words of first phrase and the rest one of user input words to generate the second phrase of Renku verse by using association method. More in details, the system searches the words associated with the above elemental words and user input word, and extract the phrases including the association words from six types of databases in the system, which are Haiku thesaurus, Kigo thesaurus, idiom thesaurus, case frame of onomatopoeia, thesaurus, and case frame.

    Furthermore, the system scores all phrases by weight: Haiku thesaurus is 3, Kigo thesaurus is 3, idiom thesaurus is 3, case frame of onomatopoeia is 3, thesaurus is 3, case frame is 1, user's relation is 5. If a phrase includes two or more association words, the system sums the weight respectively. The system chooses one of the phrases with the highest weight.
(6) After generation the second phrase of Renku verse, the similar steps to (4) and (5) are executed again. The difference is this time, the system use the elemental words from the second phrase of Renku verse and the Haiku thesaurus word to do the association.
(7) Finally, the system "hitches" these three phrases to create the Renku verse. Obviously, the difference between old Hitch Haiku system and the new one is that the new system "hitches" phrases internally and natively.

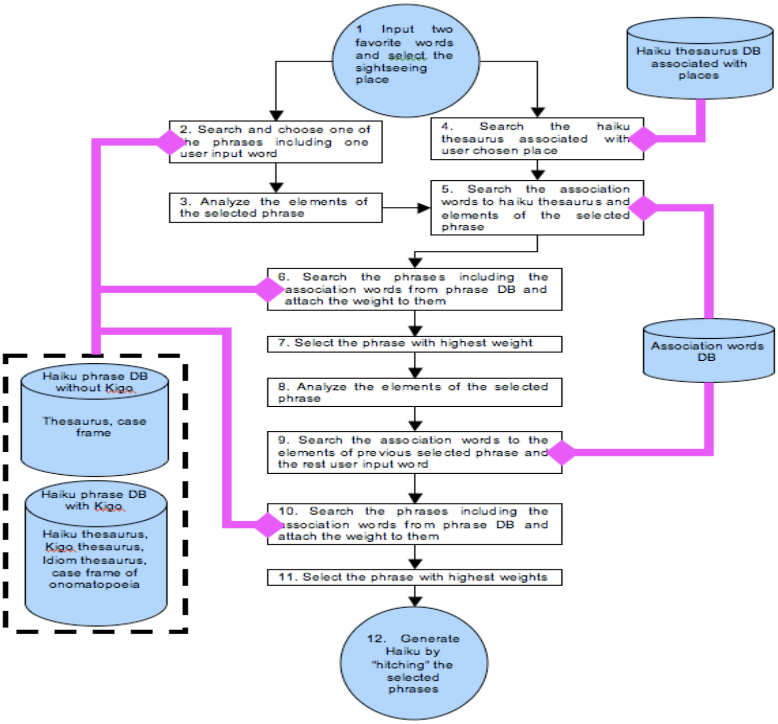


Fig. 1. Flowchart of Haiku generation algorithm

4 Database

In our new Hitch Haiku system, six types of databases, which are case frame database (about 31,000 records), thesaurus database (about 32,000 records), Haiku thesaurus database (about 2,500 records), Kigo thesaurus database (about 13,000), idiom thesaurus database (about 1,300 records) and the database of Case frame of onomatopoeia (about 8,800 records) are remained. But considering the requirements to generate Renku verse, for examples, the verse to describe the moon or flower in the specific season, we added more such information into these databases.

Furthermore, we added new Haiku thesaurus database for the sightseeing places in Kyoto city to show more local cultural background and local Haiku information which can integrate new Hitch Haiku system tightly to the sightseeing navigation system.

5 Interaction Example

With illustration figures listed below, we explain the abstract of usage of our system:

(1) A user log in the Hitch Haiku system with ID number (Fig.2).
(2) The user can select one of the functionalities provided by Hitch Haiku system by press the corresponding button (Fig. 3).

(3) The user inputs the two favorite words into text boxes and chooses the hot spot from 110 famous sightseeing places in Kyoto city (Fig. 4).

(4) After the button "Send" was pressed, the system accepts the inputs from user and shows the auto-generated Haiku phrases on a typical Japanese-style color pattern which represents the same season described in Haiku Kigo. (Fig.5).

(5) If the user does not like the auto-generated Haiku, the user can choose regeneration with same inputs by pressing "Retry" button or do the modification by himself / herself.

(6) If the user satisfies with the generated Haiku, the user can press "Ok" button, so that the system will save the result into user-note database and show the representative photo of the sightseeing place chosen by the user with address information (Fig. 6).

Fig. 2. Interface of user login

Fig. 3. Selection of functionality provided by Hitch Haiku System

Fig. 4. Input the favorite two words and choose the spot place of sightseeing place in Kyoto

Fig. 5. Generated haiku phrases on a Japanese-style picture

Fig. 6. Representative photo of user chosen place

Fig. 7. An example show the ruby for some phrase difficult to read

6 Conclusion

A new interactive Renku generation support system has been developed to help user enjoy the sightseeing in Kyoto city by using mobile phone, which is based on our previous interactive Haiku composition support system – Hitch Haiku. The theme of both researches is cultural computing. In order to get the better quality of Renku verses, we abandon the Kireji attachment step in the old Hitch Haiku system and mainly apply the association method to hitch the meaning of adjunct phrase tightly. In the other hand, to improve the cultural characteristics representation, the new Hitch Haiku system apply the Haiku thesaurus word to generate the Renku phrase so that the generated phrase can embed the feature of user chosen sightseeing place. Moreover, the new system adopts the Japanese-style color pattern as the background to show the auto-generated Renku verse to enhance the visual effect. In particular, the season that the selected color pattern represented is consentient to the Kigo of the generated Renku verse.

References

1. Tosa, N., Matsuoka, S.: ZENetic Computer: Exploring Japanese Culture. Leonardo, vol. 39(3), pp. 205–211. MIT Press, Cambridge(2006)
2. Matsuoka, S.: 1000 Books and 1000 Nights. Editorial Engineering Laboratory, `http://www.isis.ne.jp/mnn/senya/senya.html`
3. Wikipedia: Haiku, `http://en.wikipedia.org/wiki/Haiku`
4. Matsuo, B.: A Haiku Journey -Basho's Narrow Road to a Far Province-. Kodansha International Ltd. translated by Dorothy Britton (1974)
5. Lutz, T.: Stochastische text. Augenblick 4(1), 3–9 (1959)
6. Masterman, M.: Computerrized Haiku. Cybernetics, 175–183 (1971)
7. `http://digitalmedia.risd.edu/billseaman/workMajor.php`
8. Wilson, S.: Information Arts, pp. 668–671. MIT Press, Cambridge (2002)
9. Weizenbaum, J.: Eliza - a computer program for the study of natural language communication between man and machine. Communications of the ACM 9(1), 36–45 (1966)
10. Matsuoka, S.: Intelligent Editorial Engineering. Asahi Shinbun Publisher (2001)
11. Tosa, N., Obara, H., Minoh, M.: Hitch haiku: An interactive supporting system for composing haiku poem. In: Stevens, S.M., Saldamarco, S.J. (eds.) ICEC 2008. LNCS, vol. 5309, pp. 209–216. Springer, Heidelberg (2008)

Using Persuasive Technologies for Energy Consumption Management: A South African Case Study

Pieter Joubert and Sumarie Roodt

Department of Informatics, University of Pretoria
Lywnnwood, Pretoria 0002
pieter.joubertjr@up.ac.za

Abstract. In this paper the authors evaluate Chevron's Energyville, as an example of a persuasive technology, which is used to change attitudes and behaviours. The authors want to explore thissimulation in terms of it's potential to be transplanted as a persuasive technology into the South African context, given the fact that South Africa is experiencing an energy crisis currently and will continue to do so in the foreseeable future.

Keywords: Persuasive Technology, Energy, Energyville, Captology, Persuasion, Interactive Technologies, Serious Gaming.

1 Introduction

This paper investigates whether a web-based simulation, called Energyville, can be used as a persuasive technology to increase peoples' awareness of the impacts of using various sources of energy. Energyville has been designed from a developed country context but the authors want to investigate whether and how it can be used as a persuasive technology in a developing country context. Given the well documented energy crisis currently influencing South Africa, the authors have chosen this country as the basis for their research.

2 Definition of Persuasive Technology

Persuasive technology is defined as "an interactive technology that changes a person's attitudes or behaviours." (Fogg, 1998). Attitudes or behaviours concerning education, personal relationships, fitness, environmental conservation, occupational productivity, personal management & self improvement, personal finance, safety, commerce: buying & branding, preventative healthcare, disease management and community involvement/activism are some but by no means all of the attitudes and behaviours that can besubjected to persuasive technologies.

The study of interactive technologies as persuasive technology is called Captology (Fogg, 1998) and includes the design, research, and analysis of interactive computing products created for the purpose of changing people's attitudes or behaviours. According to Oinas-Kukkonen et al. (2008), the majority of persuasive technology research

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 197–203, 2009.

focuses on interactive technologies, such as computer games and mobile devices. The focus of the author's researchis on computer games as a persuasive technology.

One of the ways in which interactive technologies can function is as a medium (Fogg, 2002), where these technologies can use both interactivity and narrative to create persuasive experiences that support rehearsing a behaviour, empathizing, or exploring causal relationships. An example here is serious games because players can explore cause-and-effect relationships, as well as providing them with motivating experiences (through reward) and by practising real-world behaviour.

An example of such a persuasive technology is a virtual on-line simulation game called America's Army, which was designed to persuade American citizens to join the Army. It provides the user (person) with an idea of what it would be like to be in the Army, including the various roles (eg: medic, soldier) that are required as well as the different types of activities that take place (eg: from training to combat). A player can progress through the game by undergoing training and completing missions which allow him/her to move onto the next level (rank). There are currently over 9 million active users of the game which indicates the extent of the potential reach and impact that it can have.

Persuasive technology is based on six key principles, which are defined as follows:

- The Principle of **Cause and Effect:** Simulations can persuade people to change their attitudes or behaviours by enabling them to observe immediately the link between cause and effects (Fogg, 2002).
- The Principle of **Social Learning:** People will be more motivated to perform a target behaviour if they can use computing technology to observe others performing the target behaviour and being rewarded for it (Fogg, 2002).
- The Principle of **Suggestion:** A computing technology will have greater persuasive power if it offers suggestions at opportune moments (Fogg, 2002).
- The Principle of **Tunnelling:** Using computing technology to guide users through a process or experience provides opportunities to persuade along the way (Fogg, 2002).
- The Principle of **Virtual Rewards:** Computer simulations that reward target behaviours in a virtual world can influence people to perform the target behaviour in the real world (Fogg, 2002).
- The Principle of **Tailoring:** Information provided by computing technology will be more persuasive if it is tailored to the individual's needs, interests, personality, usage context, or other factors relevant to the individual (Fogg, 2002).

When applying these principles to America's Army, the finding is that the game can be classified as a persuasive technology as its purpose is to persuade based strongly on the principles of social learning and virtual rewards. This is because the player is rewarded during the course of the game as his/her skill level increases by being allocated to special assignments during which they have the opportunity to engage with other players to perform a specific task, which provides the social learning component.

3 Synopsis of Engergyville as a Persuasive Technology

In September 2007, Chevron Corporation and the Economist Group, launched a web-based interactive simulation game called Energyville. The purpose of Energyville is to raise awareness amongst global energy users of the impact of utilising various types of energy, namely: biomass, coal, hydro, natural gas, nuclear, petroleum, solar and wind. The impact includes:

- The Economic Impact
- The Environmental Impact
- The Security Impact

The game is based on an industrialised, developed country context where users are challenged to provide power to a city until 2030. When applying the principles of persuasive technology to Energyville the authors found the following:

- The Principle of **Cause and Effect**: Energyville does conform to this principle as it shows users the immediate impact (ie: effect) of using (ie: cause) different types of energy.
- The Principle of **Social Learning**: Energyville does not conform to this principle as users are not able to observe others and the associated rewards.
- The Principle of **Suggestion**: Energyville does not conform to this principle as the game does not provide any suggestions or hints.
- The Principle of **Tunnelling**: Energyville does conform to this principle as users are provided with useful information regarding the different sources of energy during the course of the game in order to persuade them.
- The Principle of **Virtual Rewards**: Energyville does conform to this principle as a user is allocated a point score and given a ranking.
- The Principle of **Tailoring**: Energyville does not conform to this principle as there is no mechanism for customising it to users' specific circumstances.

Based on the findings above, the authors are of the opinion that Energyville can be classified as a persuasive technology. This is because the technology, in this case being a computer game, was specifically designed for the purpose of increasing people's awareness of the impact of using various types of energy, and in doing so, hopefully persuading them to change their behaviour in terms of energy consumption.

4 The Energy Crisis in South Africa

The Energyville game, as stated above, has been tailored to a developed country context, and furthermore one which uses various futuristic technologies (in terms of the games content). South Africa, on which the authors would like to focus, as a country with a less developed energy infrastructure and technology base, is perhaps even more at risk in terms of energy issues.

The electricity supply in South Africa is managed by the para-statal company Eskom Holdings Limited, under the auspices of the Minister of Public Enterprises. Eskom uses primarily coal power stations with one nuclear power station situated at Koeberg in the Western Cape.

In the early years of the new millennium, South Africa's relatively rapid economic development, plus various other factors have resulted in a much higher demand for energy, especially in the industrial and commercial sectors. Due to this higher demand South Africa has been experiencing an energy crisis, impacting especially hard on the mining industry in South Africa, which is one of the largest industries in the country. As stated below by the Eskom website, by 2010 the energy crisis will be having a drastic impact on the South African economy and life in general:

> "South Africa's marked economic growth in recent years has propelled electricity peak demand to rise at around 4% a year in a high growth scenario (Eskom Integrated Strategic Electricity Plan). Consequently, unless something is done, by 2007 peak-period demand will exceed Eskom's ability to supply electricity during these periods, and by 2010 additional base load capacity will be required." (Eskom website).

Especially with the 2010 Soccer World Cup being held in South Africa the issue of electricity demand is a crucial priority for the country. The need to provide adequate electricity now and in the future is one of the major challenges facing South Africa in the coming years.

One attempt at alleviating the demand for energy, and thus reducing the impact of the energy crisis, is a voluntary conservation campaign. The idea behind this campaign, called Power Alert, is to provide warnings as to the consumption levels of electricity over local television stations during peak times. The belief is that alerting the population to the amount of energy they were using would result in unnecessary power usage, e.g. extra lights or appliances left on, being voluntarily curtailed. Further detail is provided by the Eskom website:

> "The principle is simple, yet effective. Over three months, a series of four colour-coded messages were broadcast during the evening peak demand period, keeping customers updated on the network status. With the colour-coding came a call to action to switch off certain categories of household appliances, depending on the state of the network (shown in Figure 1 below)." (Eskom Website).

Fig. 1. Colour-Coded Warning Messages (Adapted from Eskom Website)

This method of alerting the populace of South Africa has shown marked improvement in energy consumption and usage. One example of this is the increased awareness as shown below:

> "DSM's *(Demand Side Management)* successful marketing efforts saw a 10% increase in initial awareness levels among all target markets over 12 consecutive months." (Eksom Website) (Author's italics).

Another example is an increase in the savings of used electricity:

> "DSM management again exceeded the 2005 target, realising savings of 171 MW implemented, with energy efficiency contributing 116 MW towards the 171 MW achieved, significantly reducing greenhouse gas emissions." (Eskom Website).

5 The Energy Crisis in South Africa

The Energyville game, being a web-based flash game, is targeted at a developed context or country, where widespread internet access allows a large proportion of society to be potential impacted by the game. South Africa is generally classified as a developing country, especially in terms of its ICT (Information and Communications Technology) infrastructure. One example of this is the vast difference in internet penetration as shown by the table below:

Table 1. Internet Penetration Statistics (Adapted from Internet World Stats website)

Region	**Percentage**
South Africa	10.5 %
North America	73.6 %
Europe	48.1%

This disparity in terms of ICT can be the cause of various communications problems for the society at large in a developing country. This hinders the ability of the government to communicate with its citizens and thus impacts on its ability to affect change:

> "Finally, at the regional and global level, developing countries often have limited ability to shape trends, processes and practices that affect their economy and society." (McNamara: 2003).

This means that a persuasive technology such as Energyville, in its current format, might be limited in changing the energy habits of the South African citizenry and an adapted version is required.

6 A Persuasive Technology Proposal

TAs previously stated the Energyville game, in terms of the six principles of Persuasive Technologies, is not tailored to a South African context, both in terms of the internet capabilities (as mentioned previously) and the types of energy sources and uses in South Africa.

The authors believe that a persuasive technology, such as the Energyville game, but with appropriate changes to context and content, can be used to improve the awareness levels in South Africa to the effects of the energy crisis. Some possible context related changes could be to include new areas into the city the user has to manage that correspond more accurately with the South African context, e.g. informal settlements. Also, seeing as the simulation should focus more on energy consumption than production, the game could allow the user to adjust certain energy using habits in the virtual city and observe their effects. If the general populace were more aware of the effect that their energy saving habits could have on the electricity consumption as a whole, they will be more likely to adapt or improve their energy consumption habits. Added to this, if the game used similar visual cues, e.g. the "dials" mentioned previously, the players of the game will be able to connect the cause and effects within the game with cause and effects in real life.

Furthermore, with the high mobile phone penetration in South Africa, estimated at more than 90% of the population, a mobile based version of this game could have a much higher impact than a purely web based version. A mobile version could also have opportunities for adding Social Learning elements to the game.

7 Conclusion

In this paper the author's explore the possibility of applying a persuasive technology, more specifically an online simulation game called Energyville, which was designed from a developed world context to change attitudes and behaviours with regards to energy usage and it's three-fold impact, to help alleviate the current energy crisis being experienced in South Africa. The authors evaluated Energyville as a potential persuasive technology, and while it is sufficient for use in a developed context, the authors feel that medium will have to be tailored to the developing context in South Africa, in order to make it truly persuasive. Given the low internet penetration rates in South Africa, the authors propose using mobile technology as one possible alternative, due to the high mobile device penetration rate. This would alleviate the need for extensive training, as well as infrastructure investment, because users are already familiar with the technology and the mobile infrastructure already exists.

References

[1] America's Army, `http://www.americasarmy.com` (accessed 2008-12-07)

[2] Eskom, Energy Efficiency, `http://www.eskom.co.za/live/content.php?Item_ID=2787` (accessed 2008-12-03)

[3] Fogg, B.J.: Persuasive computers: perspectives and research directions. In: Proceedings of the CHI 1998, pp. 225–232. ACM Press, New York (1998)

[4] Fogg, B.J.: Persuasive Technology: Using Computers to Change What We Think and Do. Morgan Kaufmann, California (2002)

[5] Fogg, B.J., Eckles, D. (eds.): Mobile Persuasion: 20 Perspectives on the Future of Behavior Change. Stanford Captology Media, Stanford (2007)

[6] International Telecommunications Union, `http://www.itu.int/ITUD/ict/newslog/South+Africa+Mobile+Penetration+Edges+Towards+100.aspx` (accessed 2009-01-12)

[7] Internet World Statistics, Africa, `http://www.internetworldstats.com/stats.htm` (accessed 2009-01-08)

[8] McNamara, S.: Information And Communication Technologies, Poverty And Development: Learning From Experience. A Background Paper for the infoDev Annual Symposium December 9-10, 2003 Geneva, Switzerland (November 2003)

[9] Oinas-Kukkonen, H., Hasle, P., Harjumaa, M., Segerståhl, K., Øhrstrøm, P. (eds.): PERSUASIVE 2008. LNCS, vol. 5033. Springer, Heidelberg (2008)

[10] Stanford Captology Lab, `http://captology.stanford.edu/` (accessed 2009-01-11)

Designing Interactive Blimps as Puppets

Hideki Yoshimoto[1], Kazuhiro Jo[2], and Koichi Hori[1]

[1] Department of Aeronautics and Astronautics, University of Tokyo
yoshimoto@ailab.t.u-tokyo.ac.jp
[2] Culture Lab, Newcastle University

Abstract. In this paper we propose four models of unmanned blimps: *Robots*, *Pets*, *Agents*, and *Puppets,* according to whether they are autonomous or not and whether they are shown to people or not. *Robots* and *Pets* are autonomous and *Agents* and *Puppets* are not autonomous. *Robots* and *Agents* are shown to people and *Pets* and *Puppets* are not shown to people. Based on these models, we approach toward interactive blimps as puppets, which visualize performances from people to people with real time effects and motions. We implemented prototype applications where people could make performances through controls of the blimp's light effects and flight motions with voice via mobile phones and a physical controller. We organized observations of these prototypes at a laboratory experiment and demo exhibitions. We also discuss our models based on spectators' experience.

Keywords: Blimp, Airship, Performance, Interaction, Art, Installation.

1 Introduction

Unmanned blimps have been utilized in various fields such as hobbies, arts [1,5,7,14], science [11], and mass communications [8, 9]. Aerial Do-It-Yourself (DIY) projects are also spreading [2], which would make unmanned blimps more familiar and diversified. Blimps have following features compared with other aerial vehicles. First, blimps could fly at slow speed. Second, blimps are soft and collision between small blimps and people is not so dangerous. Third, the envelopes could be customized as canvas. These features make it easy to utilize blimps in various ways in daily life.

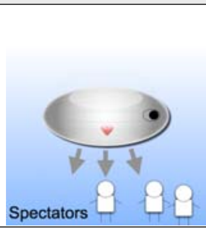
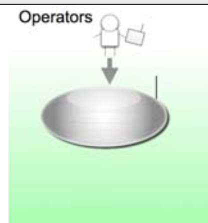
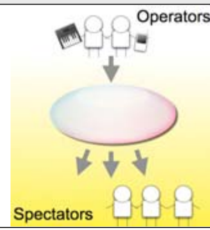
In this paper we propose four models of unmanned blimps: *Robots*, *Pets*, *Agents*, and *Puppets,* according to whether they are autonomous or not and whether they are shown to people or not. Based on these models, we approach toward interactive blimps as puppets, which visualize performances from people to people with real time effects and motions.

2 Robots, Pets, Agents, Puppets

We propose four models of unmanned blimps: *Robots*, *Pets*, *Agents*, and *Puppets*. We first divide unmanned blimps into two groups according to whether they are autonomous

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 204–209, 2009.

Table 1. Robots, Pets, Agents, Puppets

Unmanned Blimps			
Autonomous		Not Autonomous	
Not Shown to People	Shown to People	Not Shown to People	Shown to People
ROBOTS	PETS	AGENTS	PUPPETS
			

or not. Then we propose two models for each group according to whether they are shown to people or not. In these models, we define two roles of people: *Operators,* who pilot and control the blimps, and *Spectators,* who observe the blimps.

Robots. *Robots* are blimps that are autonomous and not shown to people. The blimps do their own jobs by themselves like industrial robots.

Blimps used as antennas or weather observers are classified as robots.

Pets. *Pets* are blimps that are autonomous and shown to people. Spectators observe and enjoy the blimps' flight motions as they enjoy seeing animals.

Autonomous Light Air Vessels (ALAVs) is a project where blimps are designed to flock each other or to roam for lights [1]. Blubber Bot is a DIY kit version of the ALAVs. Alan Kay's blimp in Vivarium project flies autonomously following a flash-light. Kawamura's entertainment blimp is controlled to hover and make several motions triggered by people's handclaps [6]. Ardublimp and Blimpduino are DIY autonomous blimps that fly around the ground beacons [2]. These blimps are classified as pets.

Agents. *Agents* are blimps that are not autonomous and not shown to people. Operators control the blimps for their own purposes as business clients have agencies execute their own jobs. Spectators do not concern with these activities.

"PRoP: Personal Roving Presence" by Paulos and Canny is a project where people remotely control blimps equipped with cameras and microphones through Internet to feel as if they were on board [11]. Floating Eye is an installation where people wearing head-mounted displays see wide-angle images captured by the blimp floating in the air [5]. "Naked Bandit/here, not here/white sovereign" is Knowbotic Research's installation where people control a blimp using a physical controller to attack targets with it [7]. Radio-Controlled (RC) toy blimps are controlled by operators and their purposes are the controls itself. These blimps are classified as agents.

Puppets. *Puppets* are blimps that are not autonomous and shown to people. Operators control the blimps to make performances to spectators as puppeteers animate puppets to make performances to spectators. Additional visual effects (e.g., light illuminations, graphic images, and videos) would play important roles for the performances.

Advertising or broadcasting blimps (e.g., Lightship [8]) are classified as puppets in that they are piloted and visualize representations from clients to customers. However, these representations with static images or programmed videos do not react to operators in real time. We approach toward interactive blimps as puppets that visualize performances reacting to operators in real time.

3 Designing Interactive Blimps as Puppets through Prototyping

As examples of interactive blimps as puppets, we implemented two prototype applications. One is *participative performance*, where the public control and observe the blimps. The other is *show performance*, where a skilled operator controls the blimp in front of spectators. This section explains our platform system and these applications. We also describe the observations of them.

Platform System. Our platform system consists of a blimp, wireless communication modems, and OSC-Serial Translator. We built indoor blimps equipped with three propeller motors (two for horizontal movement and one for vertical movement) and Light-Emitting Diodes (LEDs), which realize 3D motions and full-color light effects. We implemented fuzzy control systems for stabile flight motions [10]. The latest blimp is 1.15m long and the total weight is 220g. We utilized Arduino as the micro controller on board and XBee as the wireless communication modems. We employed Open Sound Control (OSC) protocol for internal communication between the platform system and applications' interfaces. OSC-Serial Translator translates the OSC messages into serial messages to communicate with the blimp.

Participative Performance. In *participative performance* the public control and observe the blimps. In our prototype application, they control the blimp's motions and

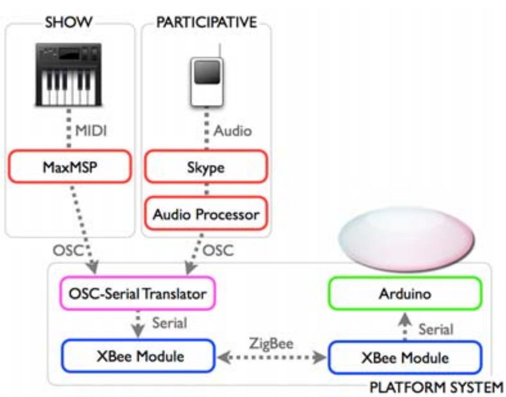


Fig. 1. Configuration

Fig. 2. Blimp

light effects with voice via mobile phones. Voice via mobile phones is advantageous as an interface for the public in that it does not require special preparations. It also enables them to control and watch the blimp simultaneously. We employed volume, pitch transitions [4], filled pauses (i.e., continuous voice like "Ahhh") [3] and word recognition as the vocal interfaces. Audio Processor analyzes incoming voice from mobile phones via Skype. The horizontal movement is controlled with combination of word recognition and the filled pauses. For example, people say, "Go, Ahhh", then the blimp goes straight during "Ahhh". The vertical movement (up/down) represents the pitch transitions of the voice (upward/downward). Brightness of the LEDs represents the volume of the voice. Color of the LEDs represents the pitch transitions. When the pitch of the voice is going up/down, the light tinges with green/red.

Show Performance. In *show performance* a skilled operator controls the blimp in front of spectators. In our prototype application, the operator control the blimp's motions and light effects with a physical controller. A physical controller is advantageous for complex controls such as those to music or dance because of the ability in numerical and prompt manipulations. We utilized sliders and buttons of a MIDI keyboard as the physical controller. Max/MSP translates the MIDI messages into OSC messages. The horizontal flight modes such as "Go Straight", "Spin", or "Right", and the vertical movement are switched with buttons. Brightness of the each color was controlled with sliders. Automatic light effects such as flash and oscillation are also programmed and switched with buttons.

Observations. We organized a laboratory experiment in a dim high-ceiling room. For observation of the participative performance application, two persons participated with their mobile phones. The vertical movement and the light effects well represented the participants' vocal controls. However, the participants had difficulties in controlling the blimp's horizontal movement with word commands because the voice through the system included noises and latency. For observation of the show performance application, one of the authors made a musical performance with the blimp. Kinds of motions such as circular flight, slow swing, and hovering and full-color light effects could represent the musical tunes. We also introduced our participative performance application in demo exhibitions at Make: Tokyo Meeting 02 (MTM 02, http://www.oreilly.co.jp/mtm/02/) and the Workshop on Media Arts, Science, and Technology (MAST) [13]. In MTM 02, more than 120 persons including children experienced controls of only light effects using voice. It was easy to learn and enjoyable for both adults and children. In MAST, more than 25 persons experienced the controls of the flight motions and light effects using voice. They enjoyed the exhibition while they had difficulties in controlling vertical and horizontal movements in balance. Light effects represented the volume and the pitch transitions. Through these observations, we recognized that verbal interfaces are not suitable for flight motions, which requires controls of multiple parameters. In both of flight motions and light effects, simpler effects work well in the participative performance and more complex effects work well in the show performance.

4 Discussion

In this section, we would like to discuss the differences between our four models based on the taxonomy proposed by Reeves et al [12]. The goal of this taxonomy is designing the spectators' experience according to how they hide or reveal manipulations and effects in interaction. Manipulations are the actions carried out by the performers (operators). Effects are the results of these manipulations. Figure 3 is the classification of unmanned blimps mentioned in the former sections.

Blimps as puppets are characterized with amplified visual effects. Manipulations in our participative performance are revealed because the public behave as both operators and spectators. Manipulations in our show performance are partially hidden because the spectators could see only rough motions of the operators. Manipulations in Lightship and Advertising Blimp are hidden because the spectators could not see the advertisers or broadcasters.

Blimps as pets are characterized with revealed effects, which means that spectators observe the blimps' flight itself but no amplified effects, and hidden or partially hidden manipulations caused by autonomous operations. Manipulations in Vivarium, Kawamura's Blimp, ALAV, and Blubber Bot are partially revealed because they operate autonomously but the spectators could affect them. Manipulations in Ardublimp and Blimpduino are hidden because they could not be affected by spectators.

Blimps as agents are characterized with transformed or hidden effects. In Toy RC Blimp and Naked Bandit, effects are transformed because the spectators could observe the blimps and the operators but they could not experience the controls that the operators experience. Manipulations of these projects are partially hidden because spectators could not see details of the controls. In Floating Eye and PRoP, effects are hidden because the spectators could not see the images that operators see. Manipulations in Floating Eye are revealed because the spectators could observe the controls of the operators. Manipulations in PRoP are hidden because the operators control the blimp remotely in distant places.

Blimps as robots are characterized with hidden effects and hidden manipulations. It is because the spectators could not see the blimps themselves.

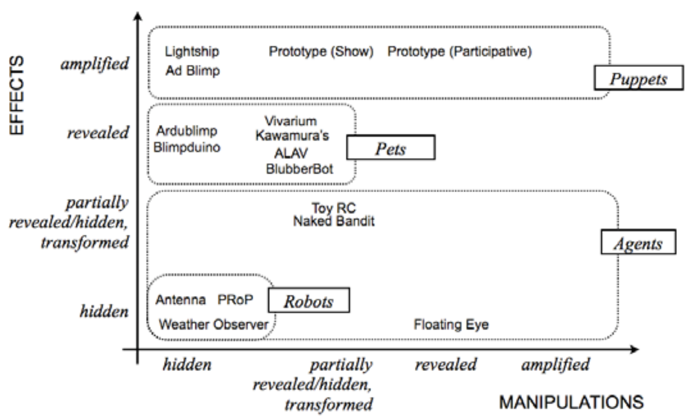


Fig. 3. Classification based on Spectators' Experience

5 Conclusion and Future Work

In this paper we proposed four models of unmanned blimps: *Robots, Pets*, *Agents*, and *Puppets*, and described our approach toward interactive blimps as puppets. Interactive blimps as puppets visualize operators' performances to spectators with real time effects and motions. As examples of interactive blimps as puppets, we implemented two prototype applications: participative performance and show performance. In the former, the public control the blimp with voice via mobile phones. In the latter, a skilled operator controls the blimp with a physical controller. In both applications the blimp visualize the performances with its full-color light effects and 3D motions. Through observations of the prototypes in a laboratory experiment and demo exhibitions, we recognized that simpler effects work well in the participative performance and more complex effects work well in the show performance. Our next interests are in interaction between multiple blimps and people. Multiple blimps could not only interact with each participant but also illuminate and direct the whole space. This kind of interaction would produce new performances in the air.

References

1. Berk, J., Mitter, N.: Autonomous Light Air Vessels (ALAVs). In: ACM MULTIMEDIA 2006, pp. 1029–1030. ACM Press, New York (2006)
2. DIY Drones, `http://diydrones.com/`
3. Goto, M., Itou, K., Hayamizu, S.: A Real-Time System Detecting Filled Pauses in Spontaneous Speech. IEICE J83-D-II, 11, 2330–2340 (2000)
4. Igarashi, T., Hughes, J.: Voice as Sound: Using Nonverbal Voice Input for Interactive Control. In: ACM UIST 2001, pp. 155–156. ACM Press, New York (2001)
5. Iwata, H.: Art and Technology in Interface Devices. In: ACM VRST 2005, pp. 1–7. ACM Press, New York (2005)
6. Kawamura, H., Kadota, H., Yamamoto, M., Takaya, T., Ohuchi, A.: Development of an Entertainment Indoor Blimp Robot Based On Hovering Control. SOFT 17, 2, 203–211 (2004) (in Japanese)
7. Knowbotic Research: Naked Bandit/here, not here/white sovereign, `http://www.krcf.org/`
8. Lightship, `http://www.lightships.com/`
9. Marvin, C.: When Old Technologies Were New. Oxford University Press, UK (1988)
10. Passino, k.M., Yurkovich, S.: Fuzzy Control. Prentice Hall, Englewood Cliffs (1998)
11. Paulos, E., Canny, J.: PRoP: Personal Roving Presence. In: ACM CHI 1998, pp. 296–303. ACM Press, New York (1998)
12. Reeves, S., Benford, S., O'Malley, C., Fraser, M.: Designing the Spectator Experience. In: ACM CHI 2005, pp. 741–750. ACM Press, New York (2005)
13. Yoshimoto, H., Hori, K.: Design of Blimps for Interactive Media and Arts. In: MAST Workshop, `http://mast.mat.ucsb.edu/`
14. Yoshimoto, H., Jo, K., Hori, K.: Design of Installation with Interactive UAVs. In: ACM ACE 2008, p. 424. ACM Press, New York (2008)

Requirements for Supporting Individual Human Creativity in the Design Domain

Uta Lösch[1], Julie Dugdale[2], and Yves Demazeau[3]

[1] Institut AIFB, Universität Karlsruhe (TH), Germany
uta.loesch@aifb.uni-karlsruhe.de
[2] LIG, Université Pierre Mendes France, Grenoble, France
Julie.Dugdale@imag.fr
[3] LIG, CNRS, Grenoble, France
Yves.Demazeau@imag.fr

Abstract. Creativity is an important activity in many professional and leisure domains. This article presents a first step towards a system which will provide a set of tools for enhancing the individual creative abilities of the user in a design task. We have identified aspects which are characterise individual creativity: motivation, domain knowledge, externalization, inspiration and analogies, and requirements handling. Based on these aspects we have defined requirements and suggest associated system functionalities.

1 Introduction

In both our personal and professional lives creativity plays an important role. Some domains, such as art, are more closely associated with creativity than others. Nevertheless, finding creative solutions to problems is a common activity in many work situations. One domain that requires a creative approach is design, for example the design of physical objects, such as mobile devices, or less tangible objects such as web sites.

Previous work on using computers for supporting creativity has largely focused on either supporting group creativity ([4], [7], [14]) or has attempted to build systems that are inherently creative themselves ([10], [12], [15]). Unfortunately, there has been little work on supporting an individual's creativity for professional tasks. Misue and Tanaka [11] focus on the early stage of the creative process and have developed a pen and paper device that allows users to express their ideas and reconfigure sketches and concepts. Kules [9] examines how search engines can be used to support creativity. Pachet [13] proposes Interactive Musical Reflective Systems to support creativity. In these systems users explore their own musical style by teaching it to the system.

Despite these attempts tools for supporting individual creativity in the task of design are rare. The ultimate goal of our work is to develop a system to support individual human creativity. In this paper we define requirements for such a system. Section 2 describes the core aspects that characterise individual creativity. Section 3 addresses the specific task of design. Section 4 discusses the desired functionalities of a support tool. Section 5 concludes with a discussion.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 210–215, 2009.

2 Supporting Individual Creativity

Creativity is an abstract notion that is hard to define. In this section we describe aspects that have been identified as being necessary preconditions for creativity.

2.1 Motivation

Motivation is a crucial aspect of creativity. It describes both the intrinsic and extrinsic reasons that lead to the engagement of the individual in a task. Intrinsic motivation occurs when an individual engages in an activity without an obvious external incentive, whereas extrinsic motivation is linked to some external incentive, such as money. Intrinsic motivation, in particular, has been shown to encourage creativity. It leads to curiosity and the readiness to take risks in deriving the solution [8]. Both of these factors increase the probability of creative results.

Intrinsic motivation is a precursor to the "flow experience" which is a mental state where a person is fully immersed in an activity and has a feeling of success in the process of that activity [3]. The flow experience is essential to the creative process and is characterised by four points: at each moment the goal of the individual engaged in the activity is clear; each action results in an immediate reaction; the difficulty of the task and the skills of the individual are balanced; focus on the task excluding all external distractions.

2.2 Domain Knowledge

Knowledge of the domain is needed during the three main phases of the creative process: the preparation phase, the idea generation phase, and the validation phase. An understanding of the problem is constructed during the preparation phase where a sound knowledge of the domain and associated skills are crucial to creativity ([2], [16]). The idea generation phase is characterised by the occurrence of insights. An insight is the recognition of a new relationship or re-organization of knowledge and leads to new ideas regarding the problem solution. The occurrence of insights is impossible without domain knowledge [8]. The validation phase concerns judging the appropriateness of the creative solution in terms of its novelty and adaptability to a situation. This requires knowledge of what already exists in the domain and the intended environment [2].

2.3 Externalisation

Externalisation is the expression of thoughts or ideas in a form outside the physical boundary of the mind, e.g., by making notes or sketches. Externalisation is a way to extend the cognitive functions of the brain [8]. It helps to understand the problem and to produce new ideas by constructing or changing the mental representation of the problem. Externalisation also helps to organise and integrate information [11] and identify missing data. In the preparation phase externalisations are the first step towards a concrete representation of the problem [5]. In the idea generation phase, possible solutions are externalised in order to produce new insights or to elaborate ideas.

3 Creativity Support in Design

Design concerns defining the characteristics of an object or procedure so that it is adapted to a set of constraints [2]. Gero describes three types of design [6]. Routine designs are a small subset of the design solution space and are found by applying "good design practice" to define further constraints. Innovative designs are obtained if some variables in the design have unusual values. Creative designs lie outside the original search space. Boden's notions of p-creativity (psychological creativity) and h-creativity (historical creativity) can be translated into this model [1]. A result involves p-creativity if it is novel to the person who created the object, whereas h-creativity results in something that is novel to society. A p-creative design lies outside of one particular designer's design space whereas a h-creative design lies outside of all designers' design spaces. Finally, s-creativity (situated creativity) occurs when the result of a creative process contains unexpected ideas and the design lies outside of the initial design space of the problem [6].

3.1 Inspiration and Analogies

Obtaining inspiration and making analogies are valuables processes in creative design [2]. Sources of inspiration give us a better understanding of the problem and can introduce additional problem constraints. Presenting sources of inspiration appears to stimulate the search for analogies. Designers who have been confronted with possible analogies during the preparation phase tend to evoke more and a greater variety of sources than designers who have not been faced with possible sources of inspiration.

3.2 Requirements Handling

Requirements define the properties that a design solution must possess. Defining additional requirements reduces the space of possible design solutions and makes the solution more precise.

There are four kinds of requirements in the design domain [2]. Explicit requirements are directly derived from the problem description and are expressed in the specification. Constructed requirements are the expression of domain knowledge. Deduced requirements are obtained by analysing current requirements in order to deduce new ones. Induced requirements are those which are introduced by the designer from his or her understanding of the problem. The difference between constructed and induced requirements is that the former are not directly related to the specification but are the expression of the designer's knowledge, whereas induced requirements have their origin in the designer's interpretation of the specification.

Some requirements are more important than others. Explicit requirements must be satisfied for each potential solution whereas preference requirements, such as colour may be less important.

4 Requirements for Creativity Support

This section suggests some functionality for a creativity support system based on the above factors.

4.1 Functionalities to Support Motivation

A system may motivate the users by addressing the four points concerning flow experience (section 2): The definition of goals and the users' awareness of them can be supported by structuring projects and defining sub-tasks. A basic way of providing immediate reactions to the users' actions is by using a WYSIWYG interface. However, more elaborate system reactions, such as having the system check if the user's input is coherent with previously defined design constraints, could be envisaged. To achieve a balance between the difficulty of the creative task and the skills of the user, users should be supported according to their level of expertise. To help users focus on their task, users should not have to deal with unrelated side activities and they should not be distracted. The system may support the first aspect by taking over routine tasks from the user, e.g. by providing templates of common designs. The second aspect is more difficult to support and could involve blocking the use of certain programs like e-mail. However, this is problematic since they may be used productively to support the design work.

Incentives or rewards provide a strong motivation for users. Different types of rewards already exist in certain systems, e.g. computer games. However, the same types of reward are not appropriate for professional systems. The issue of providing a suitable kind of reward remains an open question for us.

Finally, the system should be able to detect, resolve or prevent breakdowns when the user is stuck on a problem. This concerns the problems the user has with the creative aspect, rather than the use of the system. For example, a system could suggest appropriate design patterns to help the user overcome a breakdown.

4.2 Functionalities for Providing Domain Knowledge

Domain knowledge is used in the three main phases of the creative process. In the problem understanding phase implicit requirements and constraints that are not included in the specification must be identified. These implicit requirements often concern design principles, such as accessibility. A creativity support tool should offer possibilities of finding information concerning the specific design domain, e.g. providing a domain manual or links to interesting sources.

The idea generation phase deals with the actual design. Here, design rules that exist in the domain have to be applied. Design rules are known solutions to frequently occurring problems. The support system should be able to detect situations where the application of design rules is possible and desirable.

In the evaluation phase the applicability of the solution is judged. This is judged in terms of the requirements that have been specified. A support tool could check whether the design conforms to the pre-specified requirements.

Another issue concerns the presentation of domain knowledge. Other modes besides text, such as images, animation or sound should be considered.

4.3 Functionalities to Support Externalizations

Two aspects that are important to externalizations are the expression of ideas anywhere and anytime and interaction protocols (i.e. the way users interact with the support tool).

Since ideas can occur unexpectedly a ubiquitious note-taking functionality is very useful. A related problem is how users later find their notes.

How the user interacts with the system will greatly influence the solutions created. The user should be able: to manipulate the design object as directly as possible; to interpret the representation created in the system in different ways (these re-interpretations will help the user to be creative); to be able to annotate the design [17].

The support system should record a history of events. This addresses the idea that design solutions are progressively developed. A designer may investigate different design solutions in parallel, or return to a previous partial solution and continue work.

4.4 Functionalities Related to Providing Inspiration and Analogies

Humans find it difficult to make analogies (i.e. transfer the properties from a source to a current solution) if the source of inspiration and the object to which the property is transferred have no common properties. A creativity support tool should allow users to access a set of solutions from the same and other domains in order to provide possible sources of inspiration and realize analogies. These solutions could be organized as a dynamic graphical library which may be searched based on an ontological description of their most relevant properties. The tool should also be able to identify interesting sources of inspiration based on the currently defined requirements and the current state of the solution.

4.5 Functionalities Related to Requirements Handling

A design specification is used to derive an initial set of requirements. The creativity support tool could help in translating a specification to a set of requirements and with modifying the requirements during the design process. The support tool could also check whether the requirements are fulfilled by the proposed solution, detect any incoherencies in the set of requirements and make suggestions concerning their solution. Requirements may be interlinked, i.e. the modification of one requirement leads to the modification of other requirements. The system could be able to detect these links and to take them into account when searching for solutions to incoherencies. Since some requirements are more important than others the tool could label potential design solutions by their severity in breaking certain requirements.

4 Conclusions and Future Work

This article has discussed issues related to creativity and requirements for a tool to support individual human creativity in the domain of design. We currently develop a multi-agent system architecture that realizes the identified requirements. In the future, we will explore how the system may be personalized for different users.

References

1. Boden, M.: The Creative Mind: Myths and Mechanisms. Cardinal (1990)
2. Bonnardel, N.: Créativité et conception. Solal Editions (2006)

3. Csikszentmihalyi, M.: Creativity: Flow and the Psychology of Discovery and Invention. Harper Perennial (2006)
4. Farooq, U., et al.: Supporting creativity in distributed scientific communities. In: Proc. of the 2005 Int. Conf. on Supporting Group Work (GROUP 2005), pp. 217–226 (2005)
5. Fischer, G.: Symmetry of Ignorance, Social Creativity, and Meta-Design. Int. J. on Knowledge-Based Systems 13, 527–537 (2000)
6. Gero, J.: Computational Models of Creative Designing Based on Situated Cognition. In: Proc. of the 4th Conf. on Creativity & Cognition (C&C 2002), pp. 3–10 (2002)
7. Hermann, T.: Design Issues for Supporting Collaborative Creativity. In: Proc. of the 8th Int. Conf. on the Design of Cooperative Systems (2008)
8. Hewett, T., et al.: Creativity Support Tool Evaluation Methods and Metrics. In: Proc. of the NSF Workshop on Creativity Support Tools, pp. 10–24 (2005)
9. Kules, B.: Supporting Creativity with Search Tools. In: Proc. of the NSF Workshop on Creativity Support Tools, pp. 53–64 (2005)
10. Lirong, Q., et al.: A multi-agent system supporting creativity in conceptual design. In: Proc. of the 8th Int. Conf. on Computer-Supported Cooperative Work in Design (2003)
11. Misue, K., Tanaka, J.: A handwriting tool to support creative activities. In: Khosla, R., Howlett, R.J., Jain, L.C. (eds.) KES 2005. LNCS (LNAI), vol. 3684, pp. 423–429. Springer, Heidelberg (2005)
12. Oliveira, H., et al.: Tra-la-lyrics: An Approach to Generate Text Based On Rhythm. In: Proc. of the 4th Int. Joint Workshop on Computational Creativity (2007)
13. Pachet, F.: Enhancing Individual Creativity with Interactive Musical Reflective Systems. Musical Creativity: Current Research in Theory and Practice (2004)
14. Pinho, D., et al.: Similarity-based agents for design. In: Proc. of the 9th Int. Conf. on Computer-Supported Cooperative Work in Design, pp. 417–422 (2006)
15. Strapparava, C., et al.: Automatizing two creative functions for advertising. In: Proc. of the 4th Int. Joint Workshop on Computational Creativity (2007)
16. Warr, A., O'Neill, E.: Understanding Design as a Social Creative Process. In: Proc. of the 5th Conf. on Creativitvy & Cognition (C&C 2005), pp. 118–127 (2005)
17. Yamamoto, Y., Nakakoji, K.: Interaction Design of Tools for Fostering Creativity in the Early Stages of Information Design. Int. J. of Human-Computer Studies (IJHCS) 63(4-5), 515–535 (2005)

Sonic Gestures Applied to a Percussive Dialogue in TanGram Using Wii Remotes

Carlos D. Perales, Cristina Portalés, and Francisco Sanmartín

Universidad Politécnica de Valencia, Camino de Vera, s/n. 46022 Valencia. Spain,
carpece@doctor.upv.es, criporri@upvnet.upv.es,
frasanpi@pin.upv.es

Abstract. TanGram, is an original music score composed by Carlos D. Perales for Percussion Quartet and Nintendo Wii Remotes used as a wireless interface for Live Electronics. This paper examines how this composition explores exiting research in interaction and communication between instruments and digital sound processing discourse to produce a unique music mixed-media score, to explore the sonic possibilities of a multi-timbrical instrument with the integration of the visual gestures of an accessible device using processes of real-time gestural mimesis.

Keywords: Music, Electroacoustic, Live Electronics, Percussion.

1 Real Time Discourse Based on Digital Sound Processes

1.1 Designing Behaviour

The concept of real time processes has been extensively explored and discussed in composition and improvisation environments [1]. However the integration of wearable devices on percussion performers as an extension of them, especially in percussion pieces, opened a scope for another twist, as methodologies and research from the field of Acousmatic Music and performance explorations were introduced. To explain the integration between the wireless interface and the traditional use of mallets in percussion instruments, we start with the music score, which is designed to precise both uses.

Each percussionist has attached to his arm the Wii remote control in order to achieve a new parameter in movements and gestures (Fig. 1). Because the application runs as an organic sequence of events (DSP processes) and determines a collaborative distribution for the performers, a standalone Max MSP patch [2] was designed as a start point in the managing of processes. In order to connect the Wii dates with Max MSP a specific library object, named aka.wiiremote [3] have been used. This object allows a stable data transfer via Bluetooth. Each percussionist activate/deactivate their own events list using the 'A' button on Wii remote control.

1.2 From Gestures to Notated Sound

Although some electroacoustic composers have developed some conclusions of gestural analysis of sound for the analysis of live electronics performances, my intention was to reorient this findings towards writing a musical score that bring the freedom to

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 216–221, 2009.

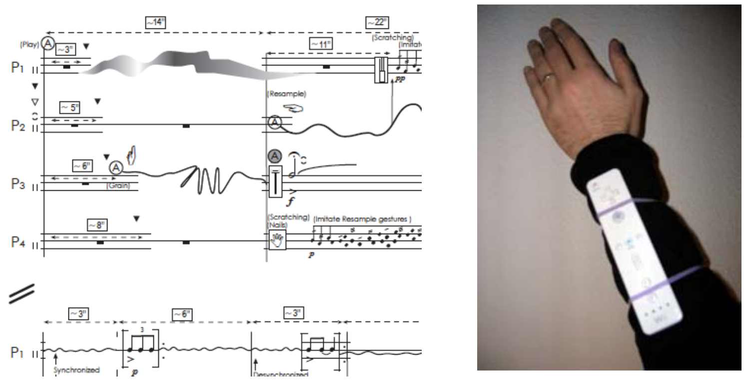


Fig. 1. Score excerpt – Wii remote

the performer to control and vary the responses to the other performers. I wanted to describe in an precise way, how after solving the tilt behaviour puzzle of acceleration in the x or y axes, ‘Kontakte Percussion Group’ [4], the performers who commissioned this electro-acoustic score, had therefore to make extensive use of their collaboration and the different performances experiences by calling on their sonic memory and creativity. After activate any process, they had to listen and interiorize their own movements for the necessary coordination between rhythms, mallets, gestures and feedback dialogue.

During this process of integration, which involved microphone experimentation and digital techniques, a sound engineer, who also produced the CD recording of the piece in winter 2009, assisted the performers.

2 Tam-Tam Morphology; A Further Step

2.1 Score Re-contextualization

This score is a re-contextualization of existing research in typology and morphology of percussion exploration and concrete sonic objects started by Pierre Schaeffer [5], continued among others by Dennis Smalley [6], and reunited by Lasse Thoresen [7]. Karlheinz Stockhausen made a specific multi-timbrical exploration of the tam-tam in his piece *Microphonie*.

Anyway I didn’t want to build a parallel world of concrete objects on the instruments but a traditional use of tam-tam with all kind of percussion beaters and two concrete sources as adhesive tape and stones curtain (Fig. 2).

The tam-tam is a very common resonator. Obviously I didn’t want to make a whole piece based on this parameter, because the electroacoustic give us the chance to explore it as an extension of it sound.

2.2 Performer Independence

Normally in electroacoustic pieces with live electronics a technician is required to control the live events and common aspects as gain, balance, etc. As the performers wanted to be

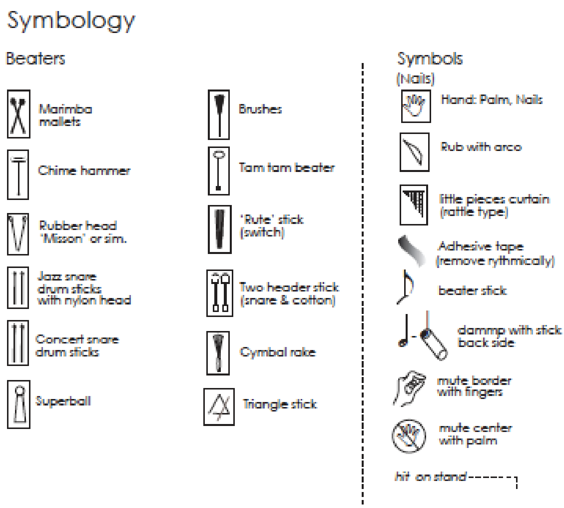


Fig. 2. Beater symbology

independents on playing the score, the patch designed is almost entirely controlled from the Wii buttons. The up and down buttons control the stereo gain. The 'A' button triggers each process saved from a coll. A matrix orient the incoming dates from each Wii to a specific DSP process (Fig. 3). Also the four microphones are activates or deactivated depending on other similar matrix reading from a coll list of events.

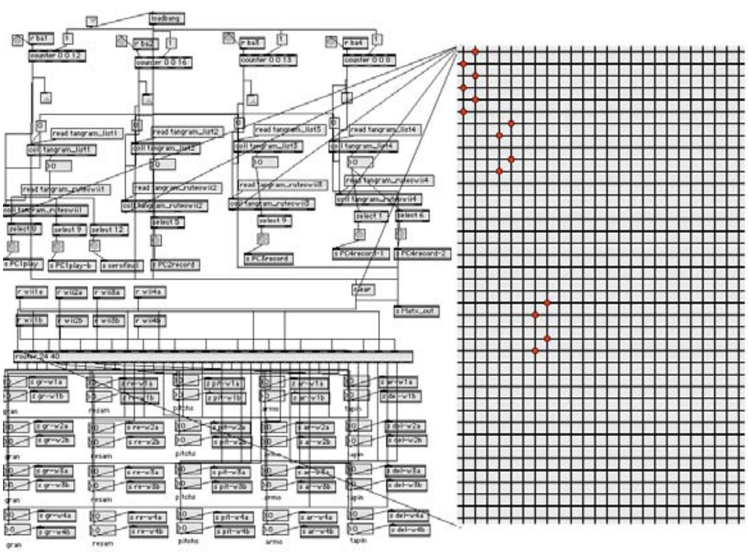


Fig. 3. Max MSP Matrix

2.3 Digital Behaviour

The musical score, including the computer part (real time processes) is a path iteration of probabilities based on Chinese TanGram game and its geometric pieces. Concepts of sound spectrum, spectral brightness, pulse, multi directionality and harmonicity act as parallel layer with the tam-tam live sound.

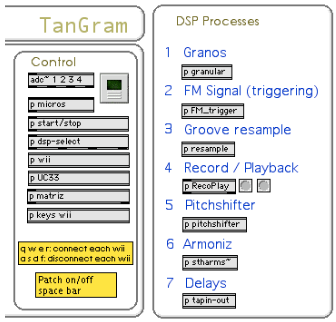


Fig. 4. DSP Processes

Seven pieces configures this easy Chinese game. Therefore, seven digital processes, seven sections and seven rhythmic motives where established for the score (Fig. 4), aiming to find beauty and musical expression in the matches and divergences between the original sound of the tam-tams and the mimetic response of the electronic discourse.

2.4 Notated Methodology

Performers need clear instructions to play around the Tam surface. A 'trigram' was established meaning the three points to hit or rub on it surface. Additional symbols were attached on score to illustrate the stand support. In order to control the damping of the tam-tam resonation, I define whether performers have to stop the resonation sharply, flatly or leave it sounding (Fig. 5). The way that they damp is a free option. In Fig. 6, the tilt of performer 1 is synchronized between the other performers.

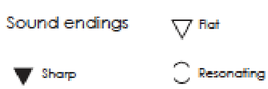


Fig. 5. Damping marks

Also damping with the left hand in two ways modifies each beater attack: by holding the edge or covering the centre area with the hand's palm. Dynamics are fortissimo (fff) to pianissimo (ppp) and some dramatic pauses are used.

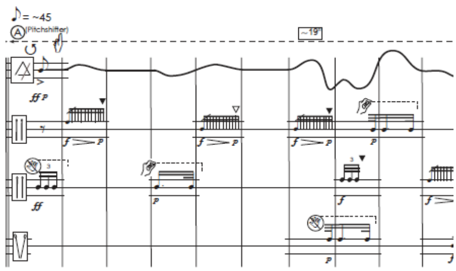


Fig. 6. Score excerpt

2.5 Rhythmic Nature

As a percussion piece the articulation and structure is entirely driven by rhythm. Only some uses of bowing and rubber techniques act as an exception of the whole piece. However, it is not just understood as rhythm in musical terms but as pace, informed by Tarkowskij's methodology [8], which is not focused on the temporal editing, but on the rhythm of the scenes (sculpting in time); in TanGram terms, it refers to the pace of different sonic scenes with a characteristic typology. By the way each electroacoustic process requires a specific tempo, so the rhythm is adapted to each response.

3 Conclusions

This paper has discussed the implementation aspects of gestural behaviour in a mixed-media piece for percussion and lives electronics called TanGram. With the introduction of the autonomy given by the Wii remote control via Max MSP patch, in this context, the author proposes a new route for creative expression, informed by Tarkowskij's ideas and methodologies in film and scene montage [9], when exposing the Ideal to its Mirror. Similarly, the score utilizes vocabulary and specific typology to precisely notate gestures and movement, generated and transformed employing computer and recording techniques. An explanation of the existing methodology to read and interpret the score leads to some detailed examples about how to notate in time concepts such as movement gestures, length variations and others, which belong to the electroacoustic vocabulary.

Acknowledgment. 'Kontakte Grup the Percussió' commissioned TanGram. The score and research behind it was realized between 2008 and 2009.

References

1. Percussion in the New-Old World. Wilfrid Mellers. The Musical Times, 133(1795), 445–447 (September 1992)
2. Puckette, M., Zicarelli, D.: Max MSP software, http://www.cycling74.com

3. Aka Universal Objects for Max MSP 4.6, `http://www.iamas.ac.jp/~aka/max/`
4. Kontakte Grup de Percussió, `http://www.kontakte-percusion.com` (last visited January 20, 2009)
5. Schaeffer, P.: Typology and Morphology of Sonic Objects. Traite des objets musicaux. Essay Interdisciplines. Nouvelle edition, Seuil (1966)
6. Smalley, D.: Spectromorphology: Explaining sound-shapes. Organised Sound 2(2), 107–126 (1997)
7. Thoresen, L.: Spectromorphological analysis of sound objects: an adaptation of Pierre Schaeffer's typomorphology. Organised Sound archive 12(2), 129–141 (2007)
8. Dempsey, M.: Andrej Tarkowskij's films. Lost Harmony: Tarkovsky's "The Mirror" and "The Stalker". Film Quarterly 35(1), 12–17 (Autumn 1981)
9. Deltcheva, R., Vlasov, E.: Back to the House II: On the Chronotopic and Ideological Reinterpretation of Lem's Solaris in Tarkovsky's Film. Russian Review 56(4), 532–549 (1997); Published by: Blackwell Publishing on behalf of The Editors and Board of Trustees of the Russian Review

TNT: Touch 'n' Tangibles on LC-Displays

Ramon Hofer and Andreas Kunz

ETH Zurich, Institute for Machine Tools and Manufacturing
Tannenstrasse 3, 8092 Zurich, Switzerland
{hofer,kunz}@inspire.ethz.ch

Abstract. In this paper, we present TNT (Touch 'n' Tangibles) - a new combination of several existing hardware technologies, which are integrated into an LC-display. TNT enables users to interact using finger touch and tangible user interfaces at the same time on an active flat panel screen, while maintaining precise identification of all interactive objects and fingers. TNT can accurately distinguish between touch and Tangible User Interfaces input by assigning different time slots to each interactive object using the same sensing technology for both methods. TNT's tracking is not affected by objects on the screen other than fingers and active Tangibles User Interfaces, which makes it ideal for use in brainstorming applications.

Keywords: HCI, Tracking, Sensor, Input Device, Multi Touch, TUI, Interaction.

1 Introduction

There already exists a wide range of multitouch detecting hardware on the market, which can be applied to small (up to 22") LC-displays. Looking into the field of research prototypes, even more technologies are available. Research in detecting multi-touch interaction reaches back to the mid 1980s, when Lee et al. [1] already designed a multi-touch tablet surface with front projection. Up to now, many new technologies were developed. One of these is DiamondTouch [2], which uses capacitive sensing of the user's fingers. SmartSkin [3] uses a similar antenna grid that can detect signal distortion on the grid by close objects. HD Touch [4] can detect reflections of touches through an LC-matrix by using a webcam and an infrared illumination. Since 2005, the FTIR (Frustrated Total Internal Reflection) multitouch detection principle [5] has inspired a lot of researchers to improve this back-projection technology or to adapt it to larger installations or other display devices such as Liquid Crystal Displays (LCD) [6].

Systems which support both – touch and TUI (Tangible User Interface) detection with reliable object identification – are not very common. Such an integrated system is Slap [7], which uses the FTIR principle on a back-projection table. It is capable of detecting touch and TUIs with attached silicone pads. Since silicone can be used to couple light out of an FTIR screen [8], it can be used as TUI identifier pads. Other projection-based technologies such as PlayAnywhere [9] use webcams above the screen to track objects. ReacTable [10] is one of the first and widely spread technologies that can simultaneously detect touch and TUI positions and orientations. Infrared

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 222–227, 2009.

light from IR (Infrared) emitters behind the interactive screen is reflected by the devices' fiducials on the screen. This allows unambiguously identifying each device. Touch can be detected as well by an interpretation of reflected blobs caused by the fingertips.

In the near future, interactive systems will most likely be based on LC-display technology. Flat screens allow applications on tables, at walls, and anywhere where the size is crucial. Thus, research increasingly focuses on interactive systems that support various methods of interactive input. ThinSight [11] is one of the first technologies that can detect touch and TUI interaction on an LC-display. The principle is similar to that of the ReacTable [10], but the screen is an LC-matrix and the camera was replaced by IR-sensors mounted in a discrete array directly behind the LC-matrix. An IR emitter is mounted next to each sensor. The emitted IR light is reflected by objects in front of the screen. Tracking update rates are around 10 Hz.

Our 'Touch and Tangibles' system (TNT) is also integrated into an LC-display. It can detect passive touch points and active TUI information (position, orientation, status) at high speed (150 Hz) by using a time-multiplexing. Additionally, passive objects other than fingers on the surface are not detected by TNT. This allows placing everyday's objects like coffee mugs, laptops, books, etc. on the screen without disturbing the tracking. We consider this a main advantage over existing systems, since we aim to support natural and intuitive group work. Applications such as collaborative sketching, brainstorming, etc, in which a high level of creativity and spontaneity is required, should not depend on the user's ability to find out the systems behaviour.

2 System Description

TNT basically consists of two technologies that are integrated into an LC-display (see Figure 1). For touch detection, an FTIR overlay is used, which acts as an IR light source. The second component is the detection array, which consists of IR sensors being mounted in a two-dimensional array behind the LC-matrix and the FTIR overlay. The detection array receives signals either from decoupled IR light of the FTIR overlay, or from synchronized active TUIs also emitting IR light.

2.1 Touch and Device Detection

An acrylic panel is mounted in front of the LC-matrix, which displays the image on the screen. IR LEDs mounted at the side of this panel couple light into the acrylic and due to the properties of refraction of acrylic and air most of the IR rays are totally internally reflected.. As soon as a finger or another "soft" coupling object is placed on the surface of the acrylic, light rays are scattered out of the acrylic and exit the overlay towards the underlying sensor array as presented in [6]. Special considerations were made concerning the sensor-sensor distance, since this significantly influences the detection performance. Here, a sensor distance of 10 mm in an overlay's distance of 15 mm was chosen. This ensures that rays decoupled by finger touches trigger sufficient sensors to perform an interpolation between the analog sensor values for the determination of an exact touch position.

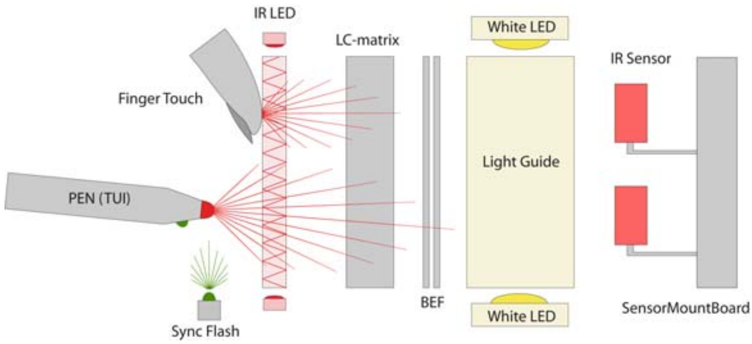


Fig. 1. Components of TNT

The device detection on TNT is basically trivial. An active device (IR light source) radiates through all the components until it hits upon the IR sensors. This cascade of components absorbs a relatively large amount of the incident IR-light. Since the radiation cone of the device's LED triggers multiple sensors at a time, the position of the device can be computed by interpolation using a predefined (measured) interpolation curve.

Two methods to distinguish between several active devices have been considered: Time- or space-multiplexed separation. If space multiplexing is used, the algorithm has to separate the sensor values into groups of potential devices. This can only be done, if the devices are not too close to each other, since otherwise (if the cones are almost overlapping) no exact assignment can be computed. This special situation can occur, if two pens are hovering and pointing at the same location. Additionally, for the space multiplexed method, the device might not be identified uniquely if removing and entering the active area does not occur at the same location. Thus, misinterpretations can easily occur.

By using the time multiplexed separation, these effects can be avoided. Each synchronized device is assigned to a separate timeslot, in which only one device is active. If the tracking is fast enough, the user will not notice the time delay for each device's tracking update. The amount of devices can easily be extended. Within TNT, a time multiplexed separation is used. A synchronization flash ensures that all the devices run in sync at any time. A slightly longer synchronization signal is used as the start frame in order to keep the logic of all devices consistent. In order to combine touch and TUI interaction using the same sensors, a timeslot for touch detection is additionally inserted into this synchronized sequence after the last TUI's timeslot. During this frame, no active device is triggered and only the FTIR-LEDs are turned on. Sensors will only receive an input from passive touch devices (fingers) on the surface in this particular timeslot.

The proposed time multiplexed detection method also permits an insertion of additional frames per TUI to transmit information such as the state or other additional data. For example, a device can be equipped with a switch that can be pressed by the user. This would alter its state to "pressed". Thus, it is possible to distinguish between a hovering (tracking) state and a pressed (dragging) state. When the TUI is removed from the screen, it remains in an "out of range" state according to [12].

3 Proof of Concept Setup and Performance

TNT is integrated into a modified 15" LC-display and uses modular sensing hardware. A SensorModule consists of two SensorControlBoards (SCB) and one SensorMountBoard (SMB). On one SMB, 128 Sensors are aligned in a 10 mm grid, resulting in an active detection area of 160 x 80 mm. In total, two of these modules are mounted behind the diffuser of an LC-display, realizing a total detection area of 160 x 160mm (see Figure 2). An SPI Bus connects all SCB to the MasterControlBoard (MCB). The MCB handles all the synchronization logic, USB data preparation, and data retrieval from all the SCB. Whereas a SCB only converts the analog sensor voltage levels into digital values and filters relevant data out of the sensor values.

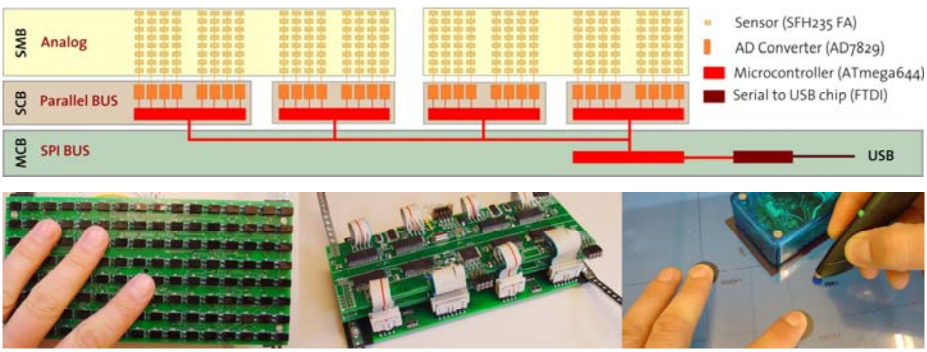


Fig. 2. Top: Schematic of two Modules (1 Module = 1 SensorMountBoard (SMB) + 2 SensorControlBoardd (SCB); Bottom: Technical proof of concept and user interaction

The FTIR touch overlay is a 3 mm thick acrylic. At every flange of the acrylic, IR SMD LEDs (880 nm wavelength) are mounted (distance of 7 mm from LED to LED). The interaction devices have an embedded microcontroller that receives the modulated synchronization pulses via an infrared sensor. The microcontroller counts the pulses and sends the corresponding bit codes via one or two (if orientation is needed) IR LEDs to the TNT system.

For the proof of concept, we have set up a small visualization software which shows graphical representations of all used TUIs together with their positions, states, and orientations. The software reads all available sensor values and computes the corresponding interpolated centers of the TUIs. For the multitouch frame, a blob detection algorithm is performed on the sensor values, which separates the touch points in different groups in order to distinguish between multiple touch points.

At the moment, the system is set to track 6 different devices (each using 2 frames) and multitouch input (one frame). This results in 13 timeslots (frames).

For the acquisition and transfer of all sensor values within one frame, approximately 0.5 ms are needed (depending on the amount of active sensors). This results in a system inherent frame rate of around 2 kHz. Depending on the amount of used frames, the effective update rate for the detection of all devices including multitouch drops to around 150 Hz.

The accuracy of touch and LED center detection is around 1 mm. The detection is very robust. Since time multiplexed detection is used, no misinterpretations occur. Touch and TUI detection can even be used concurrently. For example: Two pens in the hovering state could point at the same position and still do not affect each other's position detection.

Since an overlay is used for the multitouch detection, the offset between interaction and display layer is 3 mm, which is not distracting, but noticeable.

Image quality of the system is not as good as on the original LC-display due to the acrylic overlay and removal of two BEF (brightness enhancement films), Thus, the sensors are slightly visible behind the image display.

As with all IR-based systems, direct sunlight is another drawback of TNT and makes it impossible to reliably track any TUI or touch.

4 Conclusion and Outlook

TNT is a proof of concept that shows that touch and TUI interaction is concurrently possible on LC-displays. The combination of two known technologies shows a new way of technically combining two different interaction metaphors. It supports tracking of position, orientation and state of TUIs, while also being able to track multiple finger touches. By using the widely known FTIR method, the system is immune against detection of most objects other than finger touches.

The proof of concept is still very small (160 x 160mm). Theoretically, it could be scaled up to a 50 inches image diagonal. Then, the IR LED for the FTIR would have to be much more powerful in order to generate enough total internal reflections over the whole interactive area. On the other hand, the hardware is designed in such a way that it can be scaled up to any size although overall complexity would increase potentially and industrial manufacturing of all the PCBs would be necessary.

In the future, it is planned to add more modules to the prototype to increase the active area of tracking. Another important aspect of improvement is image quality. Different films have to be tested to find a good solution for reflecting a lot of visible light in front of the sensors but still let pass through infrared light. It is also planned to mount the synchronization flash inside the screen (behind the LC- matrix) to make the TNT even more robust and flexible. However, the tracking and detecting algorithm already works very well under normal office lighting conditions.

Acknowledgements

We want to thank the "Institute of Machine Tools and Manufacturing" for the support and Marlis Etter for helping with soldering the electronic boards.

References

1. Lee, S., Buxton, W., Smith, K.C.: A Multi-Touch three Dimensional Touch-Sensitive Tablet. In: Proceedings of the SIGCHI conference on Human factors in computing systems, vol. 15(5), pp. 21–25. ACM Press, New York (1985)

2. Dietz, P., Leigh, D.: DiamondTouch: A Multi-User Touch Technology. In: Proceedings of the 14th annual ACM symposium on User interface software and technology, Orlando, Florida, pp. 219–226. ACM Press, New York (2001)
3. Rekimoto, J.: SmartSkin: An Infrastructure for Freehand Manipulation on Interactive Surfaces. In: Proceedings of the SIGCHI conference on Human factors in computing systems, Minneapolis, Minnesota, USA, pp. 113–120. ACM Press, New York (2002)
4. Motamedi, N.: Multi-touch and Object Sensing on a High Definition LCD TV. In: CHI 2008 extended abstracts on Human factors in computing systems, Florence, Italy. ACM Press, New York (2008)
5. Han, J.: Low-Cost Multi-Touch Sensing through Frustrated Total Internal Reflection. In: Proceedings of the 18th annual ACM symposium on User interface software and technology, Seattle, WA, USA, pp. 115–118. ACM Press, New York (2005)
6. Hofer, R., Naeff, D., Kunz, A.: FLATIR: FTIR Multi-touch Detection on a Discrete Distributed Sensor Array. In: Proceedings of the 3rd international conference on Tangible and Embedded Interaction (TEI), Cambridge, UK, pp. 317–322 (2009)
7. Weiss, M., Wagner, J., Jansen, Y., Jennings, R., Khoshabeh, R., Hollan, J.D., Borchers, J.: SLAP Widgets: Bridging the Gap Between Virtual and Physical Controls on Tabletops. In: Proceedings of the 27th international Conference on Human Factors in Computing Systems, CHI 2009, Boston, MA, USA, April 4-9, pp. 481–490. ACM, New York (2009)
8. Smith, J.D., Graham, T.C.N., Holman, D., Borchers, J.: Low-Cost Malleable Surfaces with Multi-Touch Pressure Sensitivity. In: TABLETOP 2007. Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer System, Rhode Island, USA, October 10-12, pp. 205–208 (2007)
9. Wilson, A.D.: PlayAnywhere: A Compact Interactive Tabletop Projection-vision System. In: Proceedings of the 18th annual ACM symposium on User interface software and technology, Seattle, WA, USA, pp. 83–92. ACM Press, New York (2005)
10. Kaltenbrunner, M., Bencina, R.: reacTIVision: A Computer-Vision Framework for Table-Based Tangible Interaction. In: Proceedings of the 1st international Conference on Tangible and Embedded interaction, TEI 2007, Baton Rouge, Louisiana, February 15-17, pp. 69–74. ACM, New York (2007)
11. Izadi, S., Hodges, S., Butler, A., Rrustemi, A., Buxton, B.: ThinSight: Integrated Optical Multi-touch Sensing through Thin Form-factor Displays. In: Proceedings of the 2007 Workshop on Emerging Displays Technologies: Images and Beyond: the Future of Displays and Interaction, EDT 2007, San Diego, California, August 4, vol. 252, p. 6. ACM, New York (2007)
12. Buxton, W.: A three state model of graphical input. In: Diaper, D., et al. (eds.) Human-Computer Interaction - INTERACT 2009, pp. 449–456. Elsevier Science Publishers B.V. (North-Holland), Amsterdam (1990)

Entertainment Game to Support Interaction between Teachers and Students

Marcos Alexandre Rose Silva and Junia Coutinho Anacleto

Federal University of São Carlos. Washigton Luis KM 235, São Carlos, São Paulo, Brazil
{marcos_silva,junia}@dc.ufscar.br

Abstract. A narrative game is described here which main goal is to support childrens´ free expression and socialization considering their cultural background. This game can be used at school, in which students can develop a story together under the teacher's supervision. The idea is to support teachers to create characters and scenarios according to the students´ cultural context, expressed in their common sense knowledge, and consequently enabling them to get engaged on developing the story collaboratively. Also, teacher has the common sense's support to conduct the story according to the facts are being narrated, to stimuli the students' communion. This cultural sensitive RPG-like environment intends to promote a closer contact between teacher and students and among students giving them a more contextualized computer tool to be stimulated to freely express their thoughts, desires and to support them to cooperative work with teachers what is desirable for their intellectual and cognitive development.

Keywords: Collaboration, Storyteller, Narrative Game, Context, Common Sense, Education, Educational game.

1 Introduction

During childhood, according to Benford et al., [2], it is expected that young children should learn some skills to be able collaborate and realize how important is to live and to communicate with different people, because each person has his own culture, values and socio-cultural reality. These abilities are part of fundamental educational objectives because when children participate actively in their class activities, they cooperate with teacher and other students, building their own knowledge [6]. On the other hand, activities to promote work in group rarely occurs spontaneously [6] so teachers and students need to have activities and tools to leverage and support this style of studying. Thinking of that, an educational computer narrative game to help teachers to work collaboratively with their students through a cultural sensitive storytelling RPG-like environment is presented here.

2 Narrative Games

In narrative games, stories are not only understood by people; but they are also experienced in narrative processes that turn people into active characters [9], because

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 228–233, 2009.

they participate in the story, telling facts, events, and all the details necessary for the story. Fantasy in narrative games allow people, especially children, to feel safe to express themselves, to talk about situations that occur in their lives because they believe that what happens in fantasy has little or even no consequence in real life. According to Oaklander [9] children do things, behave and move in their fanciful world in the same way in their real world.

The narrative game discussed here has been developed for children from 8 to 12. According to Piaget [10] in this phase the children are in the stage called Operational Concrete Thought. In this stage child has great interest in games and finds new ways to play and to work collaboratively. They also have facilities to build and to maintain friendships using computers; they usually interact with one another virtually. These are important features for a computer game that allows people to tell stories collaboratively. In this phase children develop academic instruments such as reading, writing and basic math, and they are able to give attention. Thus, children have the capacity to read the story being told, to help to write it, i.e., to participate in building the story and to get attentive to the whole story [10]. During Operational Concrete children are willing to make friends and want to participate and interact with other children's game. Therefore, there are great chances that children can be interested in participating and interacting with the story being told collaboratively.

2.1 Contexteller

Contexteller is a narrative game inspired in Role-Playing Game – RPG [11]. In this type of game there are participants and the master, who usually is the most experienced player and his task is to present the story to the group, with characters, their characteristics, scenarios; in short, the necessary descriptions to compose an adventure with puzzles, situations and conflicts that require choices by other participants, who are the players. These players are not just spectators; they contribute actively in the story, through their characters that choose paths and take own decisions, and most of the time not foreseen by the master, contributing to the spontaneous and unexpected development of the story.

Figure 1 shows the interface available for players. This interface allows the players to see their card (I), their dice (II), and the text area (III), which allows the master to read all the messages sent to students and master during the composition of the collaborative story. In area (IV), the card, with another color and size, represents the master of the game, and area (V) shows to other characters´ card.

In this narrative game each player chooses a card that represents a character (I) and throughout his/her character the player acts, speaks, thinks and decides the character's attitudes. The card has some RPG elements, such as: Magic, Force and Experience. These elements are considered to be one of the rules existing in RPG. This rule avoids many discussions that could occur during the story.

According to Bittencourt et al., [3] and Claraparede [5], fantasy, challenges and obstacles described by the master and players during the game motivate them to play, because in narrative games, players are curious to get to know the details and are willing to participate in an entire story to achieve the goal proposed by the master. On the order hand, students can feel themselves more interest when they identify the relation between what they are seeing on the content of the story and their reality. In short, they can identify that the content is significant to their life, because it is close to

Fig. 2. The interface of the Narrative Game

their reality. Because of this, Contexteller can also help children to notice familiarity with the characters, their characteristics and plot of the story. Therefore, this gives the teacher computer support through contextualized information so that he/she can create and tell stories. This support is provided by common sense that represents cultural aspects of the students´ community. Common sense is a set of facts known by most people living in a particular culture, "covering a great part of everyday human experience, knowledge of spatial, physical, social and psychological aspects. In short, common sense is the knowledge shared by most people in a particular culture [1].

The game proposed in this paper, uses the common sense knowledge obtained by the Open Mind Common Sense Project in Brazil (OMCS-Br), developed by the Advanced Interaction Laboratory (LIA) at UFSCar in collaboration with Media Lab from Massachusetts Institute of Technology (MIT) OMCS-Br project has been collected common sense of a general public through a web site. Common sense is then processed and stored in a knowledge base as a semantic network called ConceptNet where the nodes represents concepts and they are connected through arcs that represent relations according to the Marvin Minsky's knowledge model [1]. This base intends to reflect a basic knowledge structure near human cognitive structure.

3 Setps to Create a Story at Contexteller

At the Contexteller there are six steps to support teachers to create it.

First, some information about the teacher is stored, such as: name, state, city, among others. Throughout this data it is possible to identify which teacher has created the game. Second, it is defined which students are going to participate, their names, states and school education are some important information. The teacher can identify which students played and how each student told the story. When the teacher registers his students before creating a story, he takes into consideration who are those students in order to define the characters and the plot. After the registers Contexteller gets the students´ state and filters the common sense base considering the knowledge collected from the desired profile in order to contextualize the game content for the target group.

Step three, the teacher need to choose one between two options: Creating a new game or Choosing a existing game. Step four is shown in Figure 2. In this stage the teacher needs to define six characters: one represents her/him and the others represent students. Number 6 is usually used in RPG of cards [7] and according to Díaz-Aguado [6] six is the ideal number to work collaboratively. According to Benford et al.,[2] educational research has found that working in pairs or small groups can have beneficial effects on learning and development, particularly in early years and primary education. 5 players also facilitate the teacher to monitor the whole story that is being told by the players (students). If the number of players has been greater than 5, the teacher can face difficulties in reading all the messages, in interacting appropriately during the story and in observing the development and behaviour of each character.

There are two common sense cards to support teachers to define the characters´ names and characteristics. In the first card (I), the teacher types a characteristic and searches the common sense knowledge base to obtain the characters´ names. For instance, if he/she wants to tell a story about forest and know which characters are related to the forest are taking into account, considering the student's knowledge and culture, h/she can type this place on the card. Through the common sense knowledge base the following characters can be seen: Caipora, Chico Bento, Cuca, Saci-Pererê, Iara, Curupira (from the Brazilian folklore), Big Foot, Joãozinho e Maria, Robin Hood, Elves, among others.

In the second card (II), it is possible to obtain the characters´ characteristics when the characters´ names are written on the card. For example, some characteristics coming up from Caipora´s character are: furry, short, happy, red skin, cares; she likes forest and rides a wild pig, etc. The teacher can join this information, which students know about, with the story to define characters and theirs characteristics. Step six, it is necessary to define a subject and a title for the story. In this step the teacher uses common sense through a card from which h/she can get specific information of a word typed on the card, such as: CapableOf, DefinedAs, DesireOf, MotivationOf, PartOf. For example, if the teacher is going to tell a story about forest and wants to know what students think about what exists in a forest, h/she can type "forest" on the card and select "PartOf" option, then get some data, such as: characters, characteristics, animals and other things that students believe there are in a forest.

After these steps, the game is created. Students need to identify themselves to see the title and description of the story created for them, and the teacher's name. Then, students choose a character to participate, in a similar interface shown in Figure 2, after they can change the image. This feature allows the student to express himself not only through the story but also through the image. He can choose an image that makes his character sad, joyful, angry and so on.

During the stories the teacher can also get support from common sense knowledge using a card with the same options as that of the common sense card illustrated in step 6. Figure 3 shows a situation where the teacher has described a Caipora´s attitude to protect of nature. The teacher, before typing this text, inserted on the card the word "protect" and selected the "DefinedAs" option to know what the students know about it. He/she obtained some information, such as: taking care, nurse, cherish, cradle, attend, lap, etc. This information supports the teacher during story because he/she can see some data that represent students´ language and expression and can use them to describe something. For example, the teacher can change the sentence "Caipora is protecting of the forest" by "Caipora is taking care of the forest".

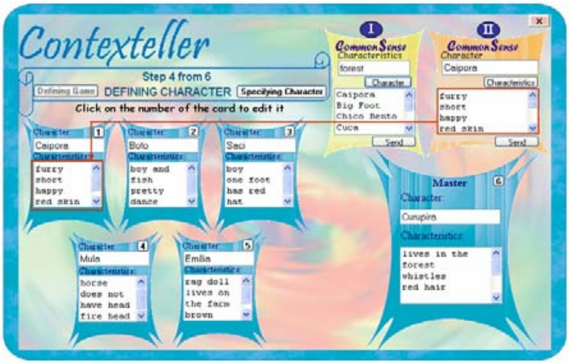


Fig. 2. Defining Characters

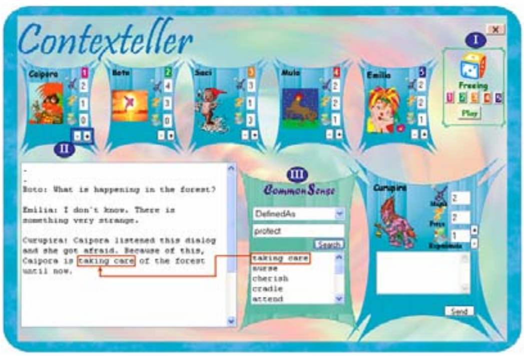


Fig. 3. Teachers´ interface

On the teacher's interface there are three different characteristics than those of students´, such as: (I) Freeing dice – allows the teacher to through his/her dice and to free the students´ dice; (II) Increasing or decreasing the experience´ value of each character; and (III) common sense card – which teacher gets information that represent students´ cultural knowledge during story.

4 Conclusions

This paper has described Contexteller, an environment for online collaborative storytelling where the players jointly develop a story under the master's supervision. The supervision is in order to guarantee that characters and scenarios are culturally contextualized, and also to promote each student's participation and space to express ideas and feeling during the story development. This game can support a teacher to interact with students which have different social and cultural backgrounds, considering these backgrounds before and during the storytelling. Contexteller, a storytelling RPG-like environment is contextualized by Common Sense knowledge in order to support teachers on defining the initial context and the content of the game's design considering students´ reality and culture. The teacher gets suggestions from the common sense to define the characters, story and its sequence. Through Contexteller, students can also learn to express, to help and to be helped because they need to tell their stories, to help their friends to achieve an objective, and to known that they also need aid to achieve their objectives.

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. (ed.) Artificial intelligence in theory and practice - WCC 2006, vol. 217, pp. 1–10. Springer, Berlin (2006)
[2] Benford, S., Bederson, B.B., Akesson, K., Bayon, V., Druin, A., Hansson, P., Hourcade, J.P., Ingram, R., Neale, H., O'Malley, C., Simsarian, K.T., Stanton, D., Sundblad, Y., Taxén, G.: Designing Storytelling Technologies to Encourage Collaboration Between Young Children. In: Conference on Human Factors in Computing Systems, pp. 556–563 (2000)
[3] Bittencourt, R.J., Giraffa, L.M.M.: A utilização dos Role-Playing Games Digitais no Processo de Ensino-Aprendizagem. Technical Reports Series, Number 031 (September 2003)
[4] Bruckman, A.: Community Support for Constructionist Learning. Computer Supported Cooperative Work: The Journal of Collaborative Computing 7, 47–86 (1998)
[5] Claraparede, E.: Funcional Education, 5th edn. SP: Comp. Nacional, 302 p. (1958)
[6] Diaz-Aguado, M.J.D.: Educação Intercultural e Aprendizagem Cooperativa. Editora Porto, Porto (2003)
[7] Fernandes, V.R.: What is RPG? RPG - Dragon Magazine Brazil, no. 123 (2008)
[8] Järvinen, A., Heliö, S., Mäyrä, F.: Communication and Community in Digital Entertainment Services. Prestudy Research Report (2002)
[9] Oaklander, V.: Windows to Our Children: A Gestalt Therapy Approach to Children and Adolescents, 335 p. Gestalt Journal Press (1988)
[10] Piaget, J.: Judgement and Reasoning in the Child, 268 p. Littlefield Adams, Richmond (1999)
[11] Silva, M.A.R., Anacleto, J.C.: A Narrative Game Culturally Contextualized by Common Sense Modeled as a Semantic Network. In: WSWEd@SBIE – Workshop on Semantic Web and Education, Fortaleza, Brazil (2008)

Multi-layer Based Authoring Tool for Digilog Book*

Jonghee Park and Woontack Woo

GIST U-VR Lab.
500-712, Gwangju, S. Korea
{jpark,wwoo}@gist.ac.kr

Abstract. In this paper, we propose multi-layer based authoring tool for Digilog Book. The main feature is that a user can author some properties of printed contents of a paper book. Those properties can be utilized for virtual contents authoring. The proposed authoring tool provides an interface to allocate some properties for printed contents. Those properties are utilized in manipulating virtual contents. As a result, users can author a realistic Digilog Book.

Keywords: augmented reality, authoring, layer, AR book, Digilog Book.

1 Introduction

Digilog Book [1] is an Augmented Reality (AR) book which provides additional information by stimulating human's five senses with multimedia contents. Digilog Book provides not only analog emotion of physical book, but also five senses experience of digital contents by combining advantages of paper books and multimedia content. The knowledge navigator [2] showed a good example in aspect of interaction between *Digilog Book* and human.

Several AR authoring tools have been proposed. Mobile Augmented Reality System (MARS) [3] introduced time line based authoring for a mobile AR. It provides Graphic User Interface (GUI) for non-programmers. If user attaches several multimedia contents to specific location in virtual space, end users can experience those contents in outdoor environments using a Head Mounted Display (HMD). Immersive authoring system for Tangible Augmented Reality (iaTAR) [4] introduced a component-based immersive authoring method. User can manipulate virtual objects by various tangible pads. However, previous works only focused on authoring properties of virtual contents without any considerations of contents in real space. Therefore, previous authoring tools are general purposed authoring tool. In addition, there are no currently authoring tools available for Digilog book from our knowledge.

In this paper, we propose a multi-layer based authoring tool for Digilog Book. The main feature is that a user can author some properties of printed contents

* This research was supported by the CTI development project of KOCCA, MCST in S.Korea.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 234–239, 2009.

of a paper book. Those properties can be utilized for virtual contents authoring. The proposed authoring tool provides an interface to allocate some properties for printed contents. As a result, virtual contents can be highly related with printed contents of a paper book.

2 Multiple Layers Based Authoring Tool

The proposed authoring tool is based on natural feature tracking method [5]. The architecture mainly consists of two components. The first one is a multi-layer based authoring system. The other one is Digilog Book viewer. In multi-layer based authoring component, it recognize a page number from camera input images using SIFT [6] features. After recognition phase, a user can start authoring tasks on the page via layer creation module. A user creates many layers by provided layer templates in layer creation module. Then, a user has to allocate some properties on each layer. The authoring information is stored in page manager module as a XML format. In Digilog Book viewer, it recognizes a page at real-time from camera captured images. After the page recognition, it loads a corresponding XML file from database and construct scene graph based on it. Finally, corresponding contents are augmented and a end user can interact with Digilog Book.

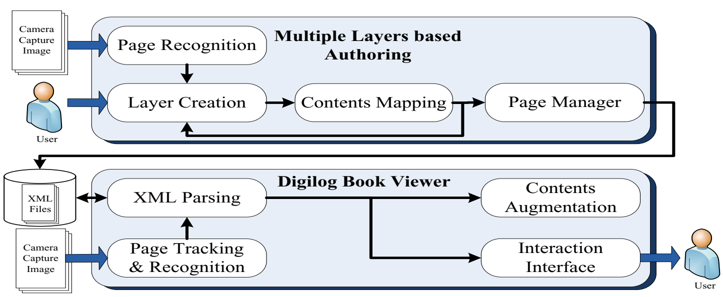


Fig. 1. Overall Architecture

Authoring tasks are accomplished simply by a repetitive process. A user generate multi-layer on each orthogonal image. For creating a layer, the authoring tool provides two types of layer templates: circle and rectangle shape. It makes creating layer task easy because a user can create a layer minimum number of control points. For more complex shape of layer, user can make a polygonal shape using a number of control points. After creating multi-layer, users have to allocate its properties such as transparent rate, layer level, and so on. As table 1 shows, layers can be categorized into two types: a contents layer and a description layer.

The properties of description layer are meaning of printed object on paper book. Description layers can be used for making relationship between virtual

Table 1. Types of Layer and its Property

Type	Detailed Type	Property
Contents Layer	Video	Play Timimg, Frequency
	Sound	
	Picture	Brightness, Chroma
	Text	Font, Size, Color
Description Layer	Description	Object

contents and printed contents. In addition, users can author multimedia contents to a specific layer. The contents layer is for augmentation of multimedia contents. The contents layer is mapped to specific contents file. The properties of each contents layer represent mapping information between contents and layer. It contains contents type oriented attributions and event which is applied in Digilog Book viewer. Figure 2 shows the concept of description multi-layer. The layer level represents a distance from a page. For example, level three layer is placed at higher position than level two layer. Therefore, a user can define occlusion between multi-layer as if multi-plane exists in a three dimensional space.

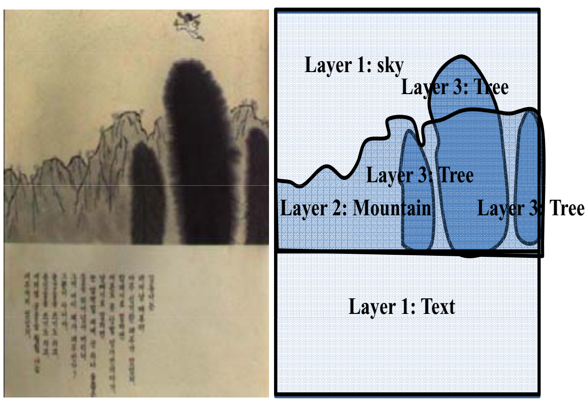
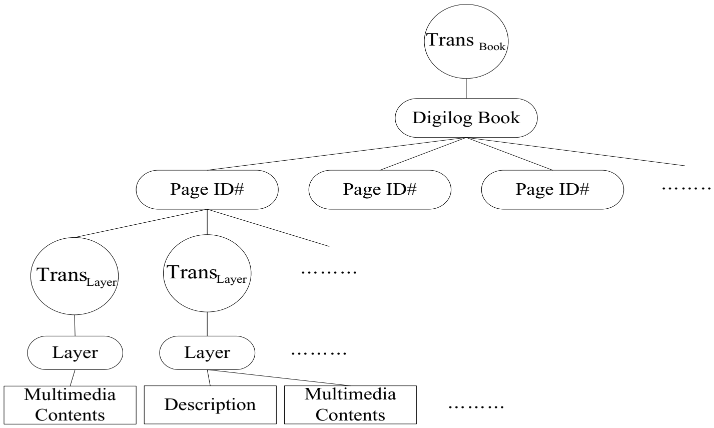


Fig. 2. Description type of multi-layer of picture area

As shown in Figure 3(a), multi-layer has a hierarchical data structure. Therefore, it is well mixed with a scene graph structure which is appropriate for 3D rendering. When a user finishes authoring one page, user has to input the page number which identifies each scene graph. The authoring information including the page number is stored as a XML form for extensibility and readability for a non-programmer user. Figure 3(b) shows an example of video contents mapped to a rectangle type of layer on page number two.



(a)

```
- <Page_Layout>
  - <Page num="2">
    - <Area type="video" shape="rectangle">
      - <Position>
          <LeftTop x="145" y="75" z="-0" />
          <RightBottom x="275" y="-22" z="1" />
        </Position>
        <Timimg>Automatically</Timimg>
        <Frequency>Repeat</Frequency>
        <FilePath>D:\cow.mov</FilePath>
      </Area>
```

(b)

Fig. 3. Authoring : (a) authoring a transparent layer with level 2 and 2D animation path with level 1 (b) its result in a Digilog Book viewer

3 Implementation and Result

We developed our system on a 2.66GHz CPU with an NVidea Quadro FX 4600 graphic card and 4 GB RAM. The proposed authoring tool is based on MFC for GUI interface. For rendering virtual contents, OpenSceneGraph (OSG) [7] is used. MSXML [8] is used for storing data.

In AR space, user's mouse input can be different according to view point because mouse input is performed in two dimensional space. Therefore, an orthogonal image of each page can be helpful for accurate user input. For that, we used reference images which are used in the page training process of markerless tracker. In addition, we provided a GUI which a user is already familiar. Figure 4 shows the user interface for the proposed authoring tool. Users can author multimedia contents in desktop environment using menu buttons on top of the main window. The main window is divided into two views. Left view shows video stream from a camera for AR view and right view shows the corresponding orthogonal image in VR space.

Fig. 4. GUI of proposed authoring tool: Left part is for an AR view and right part is for an VR view

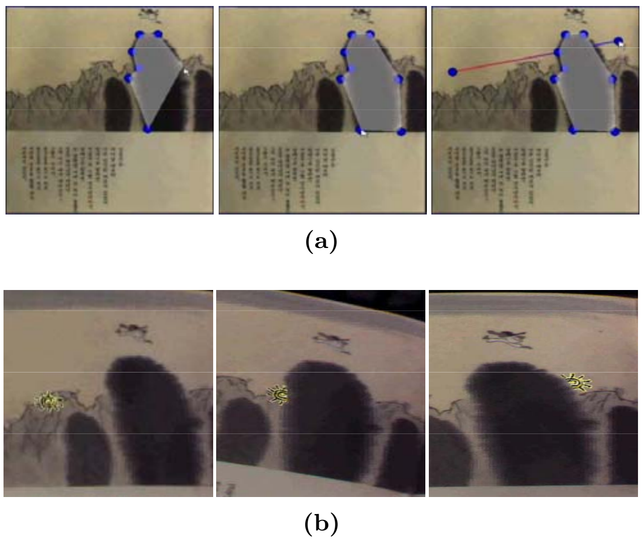


Fig. 5. An example : (a) authoring a transparent layer with level 2 and 2D animation path with level 1 in VR view (b) its result in a Digilog Book viewer

Figure 5 shows an example of multi-layer authoring. Figure 5(a) shows an authoring process in VR view. At first, an user created one layer along the printed black tree. After that, the user made a 2D animation path with a sun image behind the tree layer using a lower layer level. End users can see the Digilog Book on a monitor in desktop environment with a USB camera. It was shown as figure 5(b) in Digilog Book Viewer. The sun was seemed to pass behind the printed black tree.

4 Conclusion and Futurework

We proposed multi-layer based authoring for Digilog Book. The main functionality is that it considers contents of physical book as well as virtual contents. It is helpful to make a relationship between virtual contents and printed contents of a paper book. In implementation section, we showed an example of realistic augmentation using proposed authoring tool. As a future work, we will consider some higher level authoring for Digilog Book such as storytelling, event authoring, and haptic authoring.

References

1. Lee, Y., Ha, T., Lee, H., Kim, K., Woo, W.: Digilog book - convergence of analog book and digital content. Korea Information Processing Society 14, 186–189 (2007)
2. Dubberly, H., Mitch, D.: The Knowledge Navigator, pp. 453–469. Apple Computer, Inc. (1987)
3. Guven, S., Feiner, S.: Authoring 3D hypermedia for wearable augmented and virtual reality. In: Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings, pp. 118–126 (2003)
4. Lee, G., Nelles, C., Billinghurst, M., Kim, G.: Immersive authoring of tangible augmented reality applications. In: Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 172–181. IEEE Computer Society, Washington (2004)
5. Kim, K., Park, J., Woo, W.: Marker-less tracking for multi-layer authoring in ar books. In: International Conference on Entertainment Computing (accepted, 2009)
6. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
7. http://www.OpenSceneGraph.org
8. http://msdn.microsoft.com/enus/library/ms763742(VS.85).aspx

Community Created Narrations as Mobile Entertainment

Marjo Mäenpää, Riikka Kiljunen, and Saija Mustaniemi

University of Art and Design, Helsinki,
School of Art and Media, Pori
Siltapuistokatu 12, 28100 Pori, Finland

Abstract. In this paper we describe short mobile video story tests that have been done by using several authors/ users as story composers in the MoViE platform, which enables users to share, communicate and compose short mobile video stories in a networked community. The starting point of our study is the possibility to create an entertaining, dramaturgically intensive and coherent story from various short mobile videos composed by several authors, if there is some structure or storyline that creates the narrative. The idea was to research how separately filmed mobile videos from the same event or experience could form a dramaturgically intensive story when they are loaded into a web based story generator (MoViE). The test hypothesis was that a community – whether virtual or non virtual – could create an entertaining experience through a video montage.

Keywords: mobile media, social media, mobile community, video, experimentation narratology.

1 Introduction

The paper is based on a project where different narrative structures were tested in a matrix composed for a story generator, MoViE. In our research project we created a story generator that helps a community to compose mobile video stories following a certain structure and order. In the project video mobile phone users can present themselves as authors. Several authors can produce a common narrative with one storyline composed by the automatic story generator. The implementation is based on video database (MySql), a set of interface scripts (php) and user interface design for Nokia series 60 phones.

The project is still in process. The research group consists of Finnish resources from the University of Art and Design, Tampere University of Technology and University of Turku. The research project Mobile Social Media also has several industrial partners and collaborates with Stanford University (H-Star). The aim of Mobile Social Media –project is to search how people adopt new mobile services and how people are using mobile services in everyday life for entertainment. We wish to find out what are the things that make social, mobile media service attractive and interesting.

The rapid adoption of smart phones that integrate high quality media capture devices, coupled with trends in social networking and participatory media, provide

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 240–245, 2009.

exciting new opportunities for making video narrations and video conversations an integral component of distributed creative environments. (Multisilta, Mäenpää, 2008) According to the Statistics Finland the use of services in the mobile communication network grew in 2007 from the previous year in Finland. Call minutes from mobile phones amounted to some 13.5 billion. Similarly to the number of outgoing calls from mobile phones, the numbers of text and multimedia messages grew. The number of multimedia messages sent was nearly 29 million, which is one-third up from the year before. The use of the mobile communication network was increased also by the growth in the volume of data transmissions, which was affected by the introduction of mobile broadband connections. (Statistics Finland, 2008).

In a social media network the users typically join a community, pool or cluster sharing similar interests. The content, messages and media, such as images or videos, can be searched trough tags or keywords. The social media service may have a special interface designed for the mobile device and it can be accessed using a mobile version of the website or special client software designed for the mobile device.

It can be assumed that mobile interface users have different needs regarding the web than stationary users. The social context in mobile interfaces can derive from a need to collaborate and share information with others by keeping the interaction paths and time short.

People mostly post, share and watch entertaining videos they have shot themselves or copied from the net. Our tests have proved the hypothesis that the content in video sharing services is mostly created for entertainment purposes and that people are sharing personal videos inside a relatively closed community.

The analysis of the community that shares and composes the mobile video stories is still in progress. Personal interest and commitment towards the group or the subject is crucial in the research of social networks and studying their behavior. Commitment and some level of identification with the subject or the group helps the user/player/learner to create more easily understandable stories that are also sharable within the existing group, and as such can create a communication related situation or a story that is understandable more widely. People who share the same values or have similar interests or the same cultural background can more easily adopt and grasp even the loosest structures of a story.

People tend to see stories even where the formula of narrative, the beginning, the middle and the end, is not necessarily clearly detectable. Playing along in history writing, as well as in people's lives, there is emplotment, a perception of the narrative structure. Emplotment depends on cultural influences – stories are constructed to correspond with the prevailing and plausible form of narration of a given culture (Hietala, 2006, 94). Even a strangest story or historical anecdote becomes understandable to a person who has grown up in the sphere of western culture when it is planted into the structure of a tragedy, comedy, farce or heroic tale.

2 Research Questions and Methods

The aim of the research project is to test how different narrative structures work in digital storytelling. In the Mobile Social Media project our platform is the mobile video remix service. The research material consists of stories shot with the mobile phone video application and for the mobile phone. Therefore the narrative stories that are being created are supposed to be viewed yet again with the mobile phone. (See: Multisilta, Mäenpää, 2008).

The main research data consist of a set of test users, videos and remixes with predefined narrative structures. Methods of text and image analysis will be applied in the analysis of the remixed video data. The project will utilize and study methods of narratology, ethnography and user-centered design and technology development.

In digital media the narrative elements, place, space and time have become ubiquitous, pervasive and non-linear or fragmentary. There are several questions that have to be investigated, such as how this change has influenced the narrative situation and its elements – as typified by Genette – person, mode (or distance) and perspective. (Genette, 1988, 115).

Tzvetan Todorov coined the term narratology for the structuralist analysis of any given narrative into its constituent parts to determine their function(s) and relationships. For these purposes, the story is what is narrated as usually a chronological sequence of themes, motives and plot lines. Hence, the plot represents the logical and causal structure of a story, explaining why the events occur. (Rimmon-Kennan 1999).

Most of the structural models of narratives have been inspired by Russian Vladimir Propp (1928). Propp performed a survey of 100 fairytales in his classical study *Morfolgija skazki – Morphology of Fairytales* (Propp, 1928, 1968). The key point in his theory is that stories (fairytales) have the same general structure and therefore can easily be translated to other languages (which is, for example, much more difficult with poems). The conclusion one can make from Propp´s theory is that the narrative structures are universal and general enough to cover all human cultures i.e. the models of narrative structures lie deep in the human mind and behavior. (Greimas, 1999, 9).

The storyboard or a manuscript of an interactive story is usually composed of elements that follow (or cause) each other. Distinguished from linear narration, the actions in interactive narration might appear in an un-linear sequence, yet the story/plot follows the same form of dramaturgy. Propp´s significance lies in aiming to find a unique logic model for narrativity (same); 1) An opening of the possibility of an action – problem, enigma 2) a passage to act 3) a conclusion of the action (success, failure) (Berruti, 2005, 2/8).

Narratologists divide narration into two levels, the discourse and the story. Discourse is the surface level which can also be seen as the verbal text or the form, such as movies, poems etc. In our research case the MoViE-interphase can be seen as the discourse, preset by the community and it's users in the forms of tags and keywords. The story lies deeper and has more variable functions in the field of confrontations between the reader, viewer, user and story-composer. The story can be highly subjective and strongly culturally related. This is also why the interpretation of the story is more challenging and demands more profound knowledge on human and group behavior than the analysis of the discourse, which can in many cases be solely based on narratology.

The hypothesis that all narrations could be transferred to different media led also Barthes to make a conclusion that "…all of these narrative texts are based upon one common model, a model that causes the narrative to be recognizable as narrative…"(Ball, 2007, 175) Barthes' idea in brief was that there is a correspondence between the structure of the individual sentence and the whole text composed of various sentences. There are narratological structures in poetry and music that could be easily applied also into an interactive application.

In our project the functioning of various narrative structures was tested first with scenarios and use cases and also with the MoViE application. In the first test with the use case scenario we tested whether it is possible to compose a mobile video story – according to Propp's model – by processing morphologically varying narrative elements and dramatic action (Multisilta, Mäenpää 2008).

The scenario shows clearly the weaknesses of the hypothesis. It still remains unclear how we should let the users make decisions without previous knowledge, how to allow people to make choices according to Propp's structures without having a certain knowledge of the categories. This model makes the load too heavy for the users' competences and thus cannot be applied for the use of a large audience.

Roland Barthes was the inspirer in the second scenario of the project. Users (a group of people in a media research conference) were asked to make short mobile videos on four themes or tags – three of them based on Barthes' model – following the functions of an individual sentence, or for the actor, the subject and the object. In addition we also asked people to name videos in the given manner – e.g. place, time etc. The intention was to test how the temporal or spatial element functions in social mobile video narration.

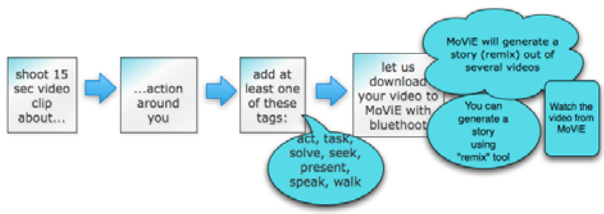


Fig. 1. Flowchart for video narrations that follow the functions of sentence. There were similar flowcharts for tagging the subject (actor, hero), object (target, problem, project) and place (location, context).

3 Analysis and Results

In the described test and the ones in process the authors themselves have analyzed the process of creating video clips by discussing the thoughts and ideas the original goal and clips gave rise to. During the analysis of the remixed clips the authors/producers were in the same secondary context. The analyzing of other tests is still in process and will be finalized by the end of 2009.

The videos were analyzed through content, structure, and dramaturgy, and compared to theoretical structures designed and presented earlier on. In several tests it has become obvious that the final product is not like a common short movie or ordinary story but, rather, montage-like.

During the 1920s the Russian filmmakers presented a hypothesis that montage is not only a phenomenon for film art, it could also appear in other forms of art. Russian Sergei Eisenstein noted that in every form of art where two elements could be linked or connected together one could also create the notion of "the third ". (Eisenstein 1964, 157).

4 Conclusion and Further Work

The test with the MoViE application has showed so far that it is possible to let several authors compose a video story or produce a common storyline shooting video clips independently from different places at different times. The story generator in application works as the *narrateé*. It composes clips together according to a certain structure that is defined in advance. However, it became evident that whatever the application, it shouldn't load too much data or advance information into the author's mind. The narrative structures in the generators' use need to be hidden or very easily adoptable by the authors.

Indeed, in our first tests we tried if it is possible to put short videos composed by several authors into one universal structure so that they automatically create a coherent story that is enjoyable and coherent, intense and whole. After several scenarios and tests with a user group we discovered that much more complex and flexible models are needed in order to create a story. Our test video montage composed under the structural model inspired by Roland Barthes had many combinations, the clips varied the themes, like movement, actor or place, and yet they created confrontation. Also our tests with MoViE shoved that to be coherent and to create suspense there has to be conflict. The object should be also an actor, the one that creates the tension.

The next phase in our project is to test the role of tagging and also to see how mobile localizing creates new elements into a story – and how they will structure the narration. The role of social media and community is also important. One of our hypothesis was about the experience of entertainment being more obvious if the community that shares the mobile videos, and in our case also composes them (several different authors), are already connected. Being connected in this case meaning that the users share the same values, same interests or same cultural background, or even the same area code, which helps them keep up their personal interest towards the mobile video stories and the network society that represents them.

Furthermore, research will be produced also from the area of community analysis. It can be seen that groups, institutions and agents are the crucial elements when it comes to creating new cultural, coherent and immaterial social patterns that can be helpful in the process of understanding and grasping the story which is produced through these mobile social platforms. Social formations may be and are included in people's understanding of and identification with a socially composed and socially functional story.

References

1. Bal, M.: Narratology, Introduction to the Theory of Narrative, 2nd edn. University of Toronto Press (1997)
2. Berruti, M.: Approaches to Narratology, lectures held at the University of Helsinki (2005)
3. Eisenstein, S.M.: Izbrannye proizvedenija v 6-I tomah. T.2 Moskva, Isskustvo (1964)
4. Genette, G.: Narrative Discourse Revisted. Cornell University, New York (translated from French, Noveau discourse du récit, 1989) (1988)
5. Greimas, A.J.: lectures given in Helsinki during 4-5 May 1979. Semiotiikan julkaisuja 1. Translated to Finnish and edited by Eero Tarasti. Yliopistopaino (1999)

6. Veijo, H.: Kertovuus. In: Ridell, S., Väliaho, P., Sihvonen, T. (eds.) Todellisuutta tarinallistamassa. Mediaa käsittämässä. Vastapaino, pp. 91–108 (2006)
7. Huttunen, T.: Montaasi ja teksti. Synteesi, 4/1997, pp. 58–79 (1997)
8. Klastrup, L.: Telling & Sharing? Understanding Mobile Stories & Future of Narratives. In: pertDAC 2007 The Future of Digital Media Culture. 7th International Digital Arts and Culture Conference, Perth, Australia, September 15-18 (2007)
9. Koskinen, I.K.: Mobile Multimedia in Action. Transactions Publishers, New Brunswick (2007)
10. Lotman Juri, M.: Merkkien maailma. Kirjoitelmia semiotiikasta (Translated from Russian by Erkki Peuranen, Paula Nieminen and Jukka Mallinen.) SN-kirjat, Helsinki (1989)
11. Multisilta, J., Mäenpää, M.: Mobile Social Media. In: Digital Interactive Media in Entertainment and Arts. In: Athens Information Technology, Athens, Greece 2008 Dimea 2008 (2008), http://hermes.ait.gr/dimea2008/
12. Propp, V.: Morfolgija skazki, Moskva (1928)
13. Propp, V.: Morphology of the Folktale. Translated by Laurence Scott, 14th edn. University of Texas Press (1998) (First English edition 1968)
14. Reponen, E., Huuskonen, P., Mihalic, K.: Primary and secondary context in mobile video communication. Pers Ubiquit Comput. 12, 281–288 (2008)
15. Rimmon-Kenan, S.: Narrative Fiction: Contemporary Poeticts. Methuen London and New York (1983)
16. Rimmon-Kenan, S.: Kertomuksen poetiikka (Translated form English by Auli Viikari) (1999). SKS Statistics Finland (2008). Volume of services in mobile networks continued to grow in 2007 (June 5, 2008), http://www.stat.fi/til/tvie/2007/tvie_2007_2008-06-05_tie_001_en.html

Hardcore Classification: Identifying Play Styles in Social Games Using Network Analysis

Ben Kirman and Shaun Lawson

Lincoln Social Computing Research Centre, University of Lincoln, LN6 7TS, UK
{bkirman,slawson}@lincoln.ac.uk

Abstract. In the social network of a web-based online game, all players are not equal. Through network analysis, we show that the community of players in a online social game is an example of a scale free small world network and that the growth of the player-base obeys a power law.

The community is centred around a minority group of “hardcore” players who define the social environment for the game, and without whom the social network would collapse. Methods are discussed for identifying this critically important subset of players automatically through analysing social behaviours within the game.

Keywords: Social Networking, Online Games, Network Analysis, Hardcore Players, Game Communities.

1 Introduction

Ludologists have known for a long time that all players of games are not equal. Among the various classifications of play styles [1] an abstract distinction is made between two classes of player: *Hardcore* players and *Casual* players. Hardcore players are defined by their high level of involvement in games, quantified by time spent in play and the scale of in-game achievements. In contrast, the Casual players are characterised by shorter, less frequent play sessions and more shallow involvement in the gaming experience[2].

1.1 Identifying the Hardcore

Essentially the Hardcore represent the pioneers of a game, and despite being a small minority of the total player-base, they help define the experience for their fellow players through their actions and behaviour.

By identifying the hardcore players and analysing play patterns it is possible to see how the game is perceived amongst these influential players. This can give vital clues to areas where the game design needs improvement. Identifying the hardcore players is not a straightforward task. Studies of gamers in the past have identified them via self-report [2] or based on the time invested in play [3].

Our approach uses Network Analysis to examine the social network within an online game and highlight the most highly connected nodes as being the hardcore centre of the game.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 246–251, 2009.

2 Network Analysis and Games

Social games are also built on top of networks - that is the network of relationships (edges in a graph) between the players (the nodes in the graph). By using network analysis methods [4] we can study the nature of social interactions within the game community. Since game interactions are recorded in real-time and network analysis is mathematical, it is possible to get a live picture of how the society within a game behaves and how it grows as time passes.

2.1 Small World Networks

Small World networks are a peculiar kind of network graph where every node can trace a path to every other node in the network[4]. They are named for the famous "Small World" experiments carried out by Travers and Milgram in the late 1960's [5] and the most familiar example of which is that of the *Six Degrees of Kevin Bacon*[6].

Social games can also expect to be Small Worlds - the larger community of players are linked through play to one another in a large contiguous social graph. It is proposed that the central nodes of a game are represented by the hardcore players - our Hardcore social gamers interact frequently with our Casual players and therefore bring them into the Small World of the game.

2.2 Analysis of a Social Game - Familiars

Familiars is a social game about collections that can be played online[1] and via mobile phone[7]. Players adopt the titular Familiars and give them free-form tasks which they would like completing. The familiar is left to be discovered by other players, who see tasks they wish to help complete. Contributions to tasks are any combination of text message, photograph and location (via GPS or self-report). Familiars was evaluated in a two month long public trial in July-August 2008.

The trial involved 157 active users and recorded 1546 distinct interactions between players. An interaction is defined as one player contributing data to the task that has been assigned to a familiar by a second player. This interaction data builds a network graph based on 157 nodes (players) and 603 distinct edges (interactions, excluding duplicates and self-interactions) that can be investigated using network analysis.

Network Properties. Based on the graph built by the Familiars players during the trial, we generated a random graph based on the same parameters of the largest contiguous social graph. N=147 due to 10 of the 157 players being totally isolated and never involved in a social interaction. By comparing the behaviour of our social players with random data we expected to see the intelligent behaviour of the players result in significant non-random patterns in the network.

[1] See http://www.familiars.eu

Table 1. Comparitive Network Properties

Property	Familiars Graph	Random Graph	KBG
Average Path Length (L)	2.314	2.584	3.65
Clustering Coefficient (γ)	0.471	0.059	0.79

Table 1 shows the properties of the Familiars graph compared with the Random graph with the same parameters. For illustration only, the values from the "Kevin Bacon Graph" (KBG) [4] are included. Average Path Length (L) represents the average length of every path from each user to each other user in the network. The higher the value of L, the more spread out the graph is, and as L decreases, it indicates how previously distant nodes have become closer via a shortcut between nearby nodes.

The Clustering Coefficient shows the average number of connections each node has within its local graph neighbourhood. For example, for every player u that is adjacent to the set of nodes V, γ is equal to the proportion of neighbours of each node v that are also adjacent to u. The clustering coefficient is the average value of γ for every node in the graph.

As the value of γ increases, the graph is more likely to break up into several, small, tightly knit graphs (i.e. a *Caveman World* [4]). A value of γ approaching 0 indicates interaction partners are selected randomly from the set of all possible players.

Compared with the random graph, the clustering of interactions in the Familiars network is striking. Since interactions are *only* between pairs, the significant difference in coefficient illustrates the intelligent and discerning nature of the players' behaviour. The high clustering points to the existence of structured groups of players within the game who are still quick to interact outside their social neighbourhood.

The Scale Free Nature of the Network. The analysis of the properties of the network confirm it as a Small World network. All social members of the network (that is, all players that interacted with someone who was not themself) appear on a single contiguous network graph, rather than several distinct groups of players. However, the relatively large amount of clustering identifies that players were discerning in their choice of interaction partner rather than random, which calls for more investigation.

For each node, we found the *Degree of Connection* (k), which is the number of nodes adjacent on the graph. In game terms this means the number of distinct individuals who have either interacted with the familiar belonging to player v, or have been the recipient of an interaction initiated by v.

Figure 1 shows the cumulative distribution of $F(k)$ against increasing values of k for both the Familiars players and the random graph with the same parameters. As can be seen, the distribution for the Familiars players strongly follows a power law. The average value of k for a network with these parameters is 8.204 but this is meaningless based on the significant difference between the random graph and the players' behaviour which shows that the Familiars network is *Scale Free* [8].

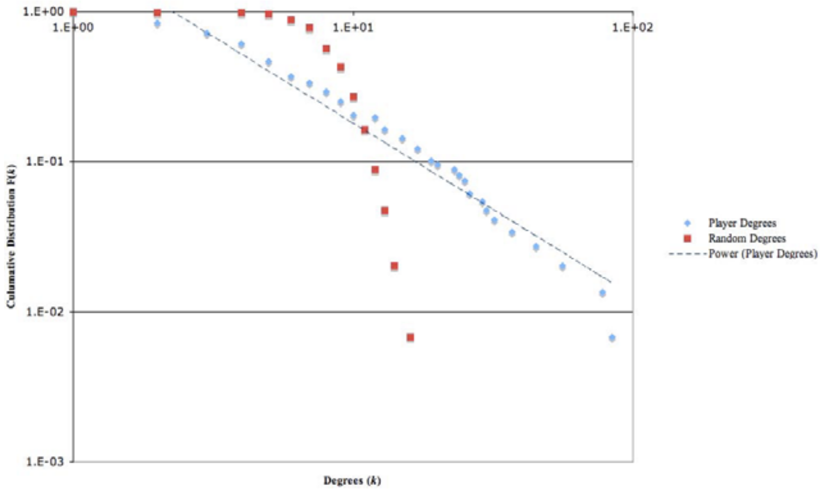


Fig. 1. Cumulative Distribution of Degrees in Familiars and Random Networks

Scale Free networks are characterised by the way that they exhibit *preferential connectivity* as they grow. In other words, as new nodes join an existing network, they are much more likely to connect to a highly connected node and therefore "the rich get richer". In Familiars this is certainly true, as new players are more likely to be involved in an interaction with a more popular player. This confirms findings in several other social applications, such as those built on Facebook [9].

2.3 Player Classification

Given the scale free nature of our typical social game network, we can classify players automatically by analysing the social attributes of a player within the social network.

Figure 2 illustrates broadly how a player base can be split into three classifications by analysing their position within the network based on "investment" or their activity within the network.

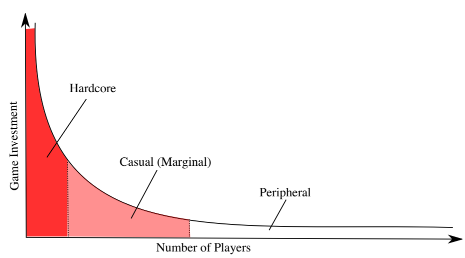


Fig. 2. Classifying the Player Base

Hardcore players are the smallest group of players of the game, but the most influential, having invested time and effort to become the most important nodes in the network of the game.

Casual players (or *Marginal* nodes in Network Analysis terms) account for the remainder of the active players who have invested a little in the game, but not as much as the hardcore.

Peripheral players are those that are only interacted with other players a handful of times at most. They are inactive and not a part of the community so appear at the very edge of the network. This kind of player accounts for the majority of nodes within the network.

Defining Classification Boundaries. There is an issue with the definition of the boundaries between the different classes of player, especially since the hardcore is such an abstract concept by definition. The boundaries can be based arbitrarily based on sudden changes in the graph [10], by proportion (e.g. boundaries at 10% and 40%) or by analysis of *cliques* (connected subgraphs) within the larger network[4].

In the case of Familiars, we analysed the network based on removing the most active players (Highest values for k) until a phase change resulted in the largest contiguous subgraph being smaller than the number of disconnected players. Practically The removed players were marked as Hardcore in that without them, the fabric of the social network fell apart. The remaining players in the disconnected graphs were identified as the marginal players and the nodes that were completely isolated from the graph were marked as the periphery (they would not even be in the game if it was not for the hardcore).

Table 2. Comparison of Player Classifications in Familiars

Classification	No. Players (%)	No. Interactions (%)	Mean k	Max k	Min k
Hardcore	18 (12.24%)	302 (50.08%)	33.56	84	17
Casual	66 (44.90%)	238 (39.47%)	7.10	15	4
Peripheral	63 (42.86%)	63 (10.45%)	2.03	4	1
All	147 (100%)	603 (100%)	8.204	84	1

Table 2 shows how the analysis through to the small world phase change split the Familiars players into the three classes. As has been demonstrated, the distribution of players and play-style is distinct - Hardcore players are involved in over half of the interactions of the game despite being just 12.4% of the community. For other social games heavy with hardcore themes, steep learning curve and high cost of entry might expect the percentage of hardcore players to be much higher than that of Familiars, when performing the same analysis.

3 Conclusions and Further Work

In this paper we proposed that the play style of player can not only be identified by their personal preference, but by the effects their play has on the social environment of the game in which they play.

Through performing network analysis of the social game Familiars, we have demonstrated that social games are likely to be Small World Scale-Free networks. The scale free nature of the networks compared with random graphs exemplifies the impact of a small but powerful group of Hardcore players who bind together the social fabric of the game community. Due to the importance of the Hardcore within the social network of a game, we demonstrated a method for classifying them, along with the marginal and peripheral players, in order that the finer-grained details of a player's activity and effects within a game world can be studied.

Our initial findings in this paper are based on a fairly small sample for a social game. The same analysis is planned to be carried out on larger scale social games (such as those on existing social networks such as Facebook) to see if the same network properties hold true.

Acknowledgements

The authors would like to thank the reviewers for their valuable feedback and suggestions for improvements to the paper. This work is part of the PASION Project, which is funded under the Presence II Initiative in the Future Emerging Technologies within the European Framework VI Programme.

References

1. Bartle, R.: Hearts, clubs, diamonds, spades: Players who suit muds (1996), http://www.mud.co.uk/richard/hcds.htm
2. Bateman, C., Boon, R.: 21st Century Game Design. Charles River Media (2006)
3. Ducheneaut, N., Yee, N., Nickell, E., Moore, R.: Building an mmo with mass appeal: A look at gameplay in world of warcraft. Games and Culture 1, 281–317 (2006)
4. Watts, D.J.: Small Worlds - The Dynamics of Networks between Order and Randomness. Princeton University Press, Princeton (1999)
5. Travers, J., Milgram, S.: An experimental study of the small world problem. Sociometry 32.4, 425–443 (1969)
6. Reynolds, P.: The oracle of bacon, http://oracleofbacon.org/
7. Kirman, B., Lawson, S., Rowland, D., Davide, F., Collovà, F., Puglia, S.: Familiars - manipulating social networks with mobile gaming. In: Proceedings of the Games Design and Technology Workshop and Conference (GDTW), Liverpool (2008)
8. Barabasi, A.L., Albert, R.: Emergence of scaling in random networks. Science 286, 509 (1999)
9. Nazir, A., Raza, S., Chuah, C.N.: Unveiling facebook: a measurement study of social network based applications. In: IMC 2008: Proceedings of the 8th ACM SIGCOMM conference on Internet measurement, pp. 43–56. ACM, New York (2008)
10. Stokman, F., Ziegler, R., Scott, J.: Networks of Corporate Power. Polity Press (1985)

Player Feedback Evaluation: Indicating Mass Public Potential for Pervasive Games

Ivo Flammer[1], Chen Yan[1], Wolf Ka[2,1], August Flammer[3], Jean-Paul Cheung[5,1], and Romain Pellerin[4,5]

[1] XiLabs, Urban Game Studio, 12 rue Vivienne, 75002 Paris, France
{ivo.flammer,chen.yan}@xilabs.fr
[2] res publica, performing arts company, 13 rue Germain Pilon, 75018 Paris, France
wolf@res-publica.fr
[3] Department of Psychology, University of Berne, Switzerland
august.flammer@psy.unibe.ch
[4] Ubidreams, 102 rue Saint-Maur, 75011 Paris, France
romain.pellerin@ubidreams.com
[5] CNAM-Cedric, 292 rue Saint-Martin, 75003 Paris, France
cheungj@cnam.fr

Abstract. Player feedback data was collected for the pervasive game "Meet Your Heartbeat Twin", an event-type LBS (Location-Based Service) game including affective computing through the player's live heartbeats. Correlation analysis of the data demonstrates broad client profile for pervasive games, covering age, gender and hobbies. The data also shows that Urban Games are clearly a novel experience; they are not an extension neither from video games, nor from mobile phone casual games. Surprisingly, the online sharing of the player's very personal data, player's location and live heart rate was not perceived as a critical issue at all. As expected, game control is crucial: to have fun, players need some adaptation time for GPS orientation and this even for a very low level complexity of mobile phone usage.

Keywords: pervasive games, ubiquitous games, urban games.

1 Introduction

Pervasive games are a new kind of games that are played in non-dedicated places and at times that can cover any moment in life. The game world combines both, virtual as well as real world objects, places and people: we define pervasive games as *games that rely on the ambiguity between real world and game universe* [1,2]. Various sensors and actuators constitute the interfaces between the real world and the game universe. These sensors and actuators are often linked through the mobile phone to internet, making the mobile phones the actual technological driver of pervasive games.

In the last decades the mobile phone has become a mass consumer product, even a necessity in human's daily life. In addition to communication, its functions now embrace location-based services, context aware services as well as entertainment. By

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 252–257, 2009.

2010, Nokia expects 50% of their telephones to be equipped with GPS (Global Positioning System) sensors [3]. This popularization of GPS, together with flat rate data billing plans will dramatically boost the number of potential pervasive game players. The only remaining limit will be the creativity of the pervasive game studios.

Although the “killer app” in pervasive games has not emerged yet, several pervasive games have had a big success: “Geocaching” [4], which allows the search of real treasures by GPS, “Botfighters” [1], which offers locating and “destroying” other players on the streets, and “Mogi-Mogi” [1], which invites to collect virtual treasures spread around in the city.

For the moment the criteria for successful pervasive games are barely known. The fact that pervasive games overlap with everyday life of the gamers, approaches their gameplay rather to strategy games than to action games. Pervasive game scenarios have to deal with unpredictable individual and social behavior, with simultaneous involvement of the player in several tasks, some in the game, others related to his daily life [5,6].

So as we glimpse into the near future, the following question arises: who will be the players of the pervasive games? This paper reports an empirical study on the player profiles of the pervasive game “Meet Your Heartbeat Twin” [7]. The results open a promising perspective on a potential mass public market for pervasive games.

2 Description of the Game “Meet Your Heartbeat Twin”

A typical “Meet Your Heartbeat Twin” (MYHT) session comprises around 10 players who are equipped with heart rate sensors and GPS. A game session lasts around 20 minutes. Every participant sees a geographical map of the game area on his or her mobile phone that indicates the positions of other participants with the same heart rate as the holder of the mobile phone. The positions of the participants with different heart rates are invisible to the player. The aim of the game is to physically meet a person with the same heart rate. The challenge is that by approaching a heart rate twin, one’s heartbeat might change either due to physical effort or to emotional arousal. New players will appear on the player’s map because his heartbeat now matches theirs.

Game spectators can follow the game session on a giant screen, on which the game map with the totality of the players’ data is shown. Thus, the spectators have full and live access to the emotional and geographical dynamics of the game. Once the game is over, a next group of around 10 persons takes turn.

Concerning the technology of MYHT: Bluetooth connects the sensors to the respective mobile phone. Client and server code are written in JAVA using the mobile phone game middleware GASP [8].

3 Data Collection

Data collection for this study was done during two different events. The first event was the digital art festival “LeCube” that was held in Issy-les-Moulineau/France in

July 2008 [9]. In order to solicit common citizens to participate at the projects, the festival was free of charge, mainly held outdoors and spread out over the city center. 15 game sessions of MYHT were played.

The second event was a popular neighborhood gathering in Paris. A citizen association that organizes regularly such events in the 2nd arrondissement in Paris, invited us to present MYHT and install a game counter in July 2008. MYHT was announced in the local press [10], 3 game sessions where played.

We asked the players to fill out the questionnaire after they had participated at the game session and returned their equipment. About 80% of all players were willing to spend about 5 minutes to answer the 17 questions, i.e., 83 in the first event and 13 in the second event. The number of 96 responding participants ensures strong power of the statistical analysis.

The multiple-choice questionnaire consisted of two types of questions; the ones related to the player's profile and the others related to the player's opinion about the game. In the next section, we will present the frequency distributions of the players' parameters (profile and opinions) and statistically analyze the relations between them. All statistical data analysis was done by SPSS. The level of statistical significance was defined as $\alpha = 0.05$. This means that the error probability is smaller than 5 percent. In cases where we emphasize the result of no-difference or no-correlation, we required $\alpha = 0.10$.

4 Results

We handle the data from the two events in one single data pool. Where we report numbers, we present percentages, but all statistical analyses were done with actual frequencies.

4.1 Player Profiles

The available data comprises age, gender, residence, interests, and GPS experience.

In terms of age, the players were between 12 and 62 (Figure 1) with a median of 25 years.

It may not surprise that the game attracted primarily younger participants. But it is especially noteworthy how many participants of the second half of the life-span were attracted by the game too.

The gender distribution of the players was near to even, with a slight preponderance of female participants: 55% women vs. 45% men.

In order to identify the place of residence, the player's postal code was asked for. For further statistical interpretation, however, only two categories where retained: urban with 52% and sub-urban with 48% of the players.

Two questionnaire items covered the players' interests, i.e., their preferred activities or hobbies and the type of games they usually play (Table 1). For both questions, several choices where proposed, each choice could be answered by "yes" or "no".

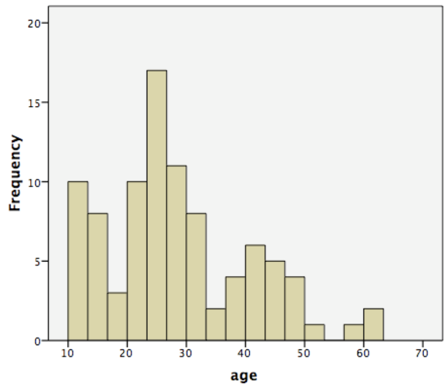


Fig. 1. Age distribution of participants

Table 1. Interest: (a) Preferred activities, (b) Games usually played

Preferred activities		Games usually played	
movies	65%		
theatre	43%	social games	41%
sports	30%	console games	43%
strolls	45%	mobile phone casual games	14%

Table 1b shows that social games and console games are more frequent than the more modern mobile phone casual games. It is remarkable that our MYHT game attracted many participants who by far do not appear as game freaks. Pervasive games seem to offer something that also attracts new segments of game consumers.

Geographical orientation skills and prior use of GPS help the gamers to understand the graphical interface and the control of the avatar position. This is why we added a player profile question in order to distinguish prior GPS users from novices. 14% of the players had previous regular GPS experience, 86% did not.

4.2 Player Opinions

Most of the opinion questions offered Likert scales allowing a choice among 4 degrees of agreement with a given statement.

As Figure 2a shows that 91% of the participants liked the game very much (70%) or rather much (21%), and only 1 participant did not like it at all.

In order to find out whether the players of a certain group like the game more than some other group, we analyzed the correlation between the opinion answers and player profile.

Age: In order to enhance the power of the analysis, we compared those of age 25 or younger with those of age 25 or older. The analysis yielded no statistical significance ($t = 1.15$; $df = 60$; $p = 0.25$). An additional analysis with all participants also yielded a non-significant correlation between liking and age ($r = -0.07$; $p = 0.52$). This is an

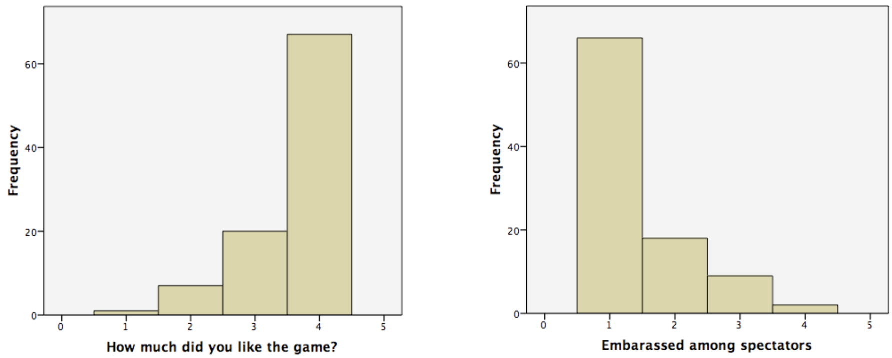


Fig. 2. a) Frequency distribution to the question “How much did you like this game?”. b) Frequency distribution to the question “Did you feel embarrassed to play among non-participant spectators?”

important finding: Pervasive games are attractive to both, young and older users. A similar game, “GO!”, by XiLabs and Visual System was staged in a public park in Shanghai and we found equal success for young university students during the weekends as retired seniors during morning weekdays.

Gender: There was no significant difference between female and male players ($t = 0.62$; $df = 90$; $p = 0.53$). Apparently, men and women like pervasive games equally well. As with MMOGs [11] and social games [12], it seems that pervasive games find its prospective clients equally in men as in women.

The hobby preferences did not influence the liking of MYHT either (all p’s > 0.50). This finding indicates the independence of pervasive games as a new type of game: they are neither video games on mobile phones, nor sophisticated casual mobile phone games.

How difficult was this game to play? 75% of the players found the game “very easy” (18%) or “rather easy” (57%). Players with prior GPS experience surprisingly did not find the game easier than players without prior GPS experience. However, comparing mobile phone game users with those not used to mobile phone games yielded a sizeable difference in easiness, i.e., 3.23 vs. 2.85 (t[one-sided] = 1.82; $df = 92$; $p = 0.04$). Although MYHT game interface was set up to avoid any difficult telephone manipulation: playing uses a single button, it still seems to need some training to be played by anyone.

As the game was played in public and under the eyes of non-playing spectators, we asked the players whether this made their participation uncomfortable. According to Figure 2b, 70% of the players felt “not at all” uncomfortable. This was equally valid for both genders, both age groups, and for mobile phone gamers vs. mobile phone non-gamers (no significant differences, $p > 0.10$ in each case).

There could be another reason for feeling uncomfortable, namely publicly sharing a personal attributes like the heart rate. Data analysis suggests that this is not the case, the answers to a further question showed that 84% of the players felt "not at all" uncomfortable.

5 Conclusions

MYHT investigation shows that pervasive games are clearly a novel experience; they are neither an extension from video games, nor an extension from mobile phone casual games. With the pervasion of 3G/GPS into the general public, pervasive games are being transformed from elite media into mass media, with possible attractiveness to broad player profile, including age, gender and hobbies. The extension of the magic circle is a crucial feature of pervasive games [13]. It seems that it is also the very same characteristic that makes it attractive to players: players love to play in the middle of their city, in the midst of non-players.

References

1. Yan, C.: Adaptive Multiplayer Ubiquitous Games: design principles and an implementation framework. Joint research PhD thesis with Orange Labs and CNAM, Paris, France (2007)
2. Flammer, I.: Urban Games. In: Taipei SIGGRAPH proceedings (2008)
3. Auchard, E.: Nokia sees half of cell phones with GPS in 2010-12. Reuters-Technology News (May 14, 2008)
4. Geocaching, `http://www.geocaching.com`
5. Tamminen, S., Oulasvirta, A., Toiskallio, K., Kankainen, A.: Understanding mobile contexts. In: Chittaro, L. (ed.) Mobile HCI 2003. LNCS, vol. 2795, pp. 17–31. Springer, Heidelberg (2003)
6. Iacucci, G., Mäkelä, A., Ranta, M., Mäntylä, M.: Visualizing Context, Mobility and Group Interaction: Role Games to Design Product Concepts for Mobile Communication. In: Proceeding of COOP 2000, Designing Cooperative Systems Conference, pp. 53–65. IOS Press, Amsterdam (2000)
7. Flammer, I., Ka, W., Skraba, R.: Meet Your Heartbeat Twin. In: Proceedings of 4th International Symposium on Pervasive Gaming Applications. PerGames 2007, pp. 157–158 (2009)
8. Pellerin, R., Gressier-Soudan, E., Simatic, M.: uGASP: an OSGi based middleware enabling ubiquitous multiplayer gaming. In: ICPS 2008, Sorrento, Italy (2008)
9. Beaux Arts Magazine, "Le Cube Festival" (June 2008)
10. Un portable pour rencontrer l'âme coeur, Le Parisien (July 19, 2008)
11. Parks Associates: Electronic Gaming in the digital home. Edited by Yuanzhe Cai, Director of Broadband and Gaming (September 2006)
12. Canary, D.J., Emmers-Sommer, T.M., Faulkner, S.: Sex and Gender Differences in Personal Relationships. Guilford Press, New York (1997)
13. Montola, M., Waern, A., Nieuxwdorp, E.: Domain of Pervasive Gaming, IperG, Work-Package WP5, Deliverable D5.3B (2006)

A Real-Time Video Illustration Using CUDA

JiHyung Lee[1], Yoon-Seok Choi[1], Bon-Ki Koo[1], and Chi Jung Hwang[2]

[1] Electronics and Telecommunications Research Institute,
138 Gajeong-ro, Yuseong-gu, Daejeon, 305-700, Republic of Korea
{ijihyung,ys-choi,bkkoo}@etri.re.kr
[2] Chungnam National University, Department of Computer Science,
79 Daehangno, Yuseong-gu, Daejeon, 305-764, Republic of Korea
cjhwang@cnu.ac.kr

Abstract. According to advancements in video technology, there are lots of needs for various special effects of videos. The conventional image-transform effects could be applied to video streams, but non-photorealistic rendering effects are not easy to apply. For example, cartoon or illustration effects have expensive costs in video transformation which makes it difficult to execute in real-time. In this paper, we suggest a video transformation system with illustration effects. It is designed to apply the illustration effects to the video stream directly and is implemented to achieve real time performances using the GPU hardware with NVIDIA's CUDA.

Keywords: non-photorealistic rendering, video, illustration, real-time, CUDA.

1 Introduction

Recently, videos have become quite common in a human's life. That is, videos are a very familiar and popular media. So, there are demands to make videos with unique style.

The easiest way to make a unique video is putting in special effects in it. To achieve this goal, many special effects for videos were designed in the past. The early results are derived from image transformations. The simple color conversions such as black &white and sepia toning can be easily applied to videos on a real-time through conventional applications.

However, video effects from more complex image transformations like non-photorealistic rendering effects can't be applied easily. Non-photorealistic rendering effects are a variety of effects as if a person is directly painting a picture on the simple image. For example, illustration and watercolor effects in an image takes a long time to be generated and it is similar in videos. Several studies have been made on video effects in the fields of non-photorealistic rendering, but those were simply used as small images.

This paper aims to apply non-photorealistic rendering effects to various image sizes in videos in real-time. Among non-photorealistic rendering effects, illustration effects are emphasized. For real-time performances, it is designed and implemented using the GPU hardware with NVIDIA's CUDA.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 258–263, 2009.

2 Previous Work

Gooch et al. [1] suggests a facial illustration method. In his research, the illustration technique and the caricature system, which are put in features of an individual, are introduced. Even though he is mainly dealing with the caricature system, his illustration technique for the face description also shows positive results.

Holger [2] introduces the technique which converted images in videos into images of cartoon style on real-time. First, an image is abstracted by applying the bilateral filter repeatedly. Second, the luminance quantization and DoG edge detection technique are applied to the abstracted image. Results of the processes are combined and then the final cartoon-style image is produced. And Klein [3] suggests the video mosaic method which uses not only simple image but also video stream.

3 Video Illustration

In this chapter, our real-time video illustration is described. At first, we mention basic illustration algorithm and explain modified algorithm for real-time processing. Then, we deal with the NVIDIA's CUDA code in order to implement our algorithm.

3.1 Illustration

Illustration effects are based on brightness perception of black and white images. The brightness of an object depends on the light reflected by itself and the object's background. Even if the brightness is a little different, relatively, it can be seen the same. As shown in Fig. 1, below four circles have equal intensities. While the circles in (a) seems to be the same, those in (b) seems to be different due to their backgrounds.

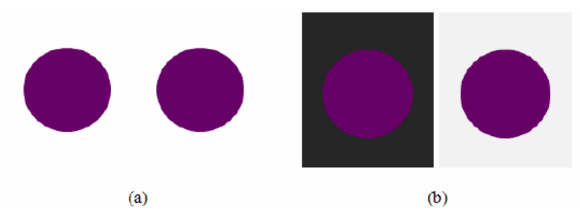


Fig. 1. Example of brightness perception

The illustration is composed of three basic operations: differentiation, integration, and threshold. First, two blurred images can be generated by applying two different sizes of Gaussian blur filters to an original image. Second, the difference between the corresponding pixels of two images is calculated and the value is integrated. The size of the Gaussian blur filter increased than the former step with 1.6-fold, and the same task is repeated several times. Finally, the last image based on an original image value, an accumulated difference value, and a pre-defined threshold value is produced. Refer to the below algorithm in Fig. 2.

```
Program Illustration(SrcImage)
{
  GrayImage = ConvertToGray(SrcImage);
  v1, v2, b : Image ;
  for s=1 to S {
    fAlpha = pow (1.6, s);
    nKernelSize = DecideKernelSize(fAlpha);
    GaussianFilterWeight= MakeGaussianFilter(nKernelSize);
    v2 = GaussianFilter(v1, GaussianFilterWeight);
    b +=  (v1 - v2) / (coeff + v1);
    v1 = Copy(v2);
  }
  FinalImage = Threshold(b, GreyImage, fThresValue);
  return FinalImage;
}
```

Fig. 2. Image illustration algorithm

3.2 Separable Gaussian Blur Filter

We divided the Gaussian Blur process into two passes to reduce computation costs. In the first pass, a one-dimensional kernel is used to blur an image in only horizontal or vertical direction. In the second pass, another one-dimensional kernel is used to blur in remaining direction. The results of those passes are the same as convolving with a two-dimensional kernel in a single pass. Equation 1 shows the basic Gaussian Blur filter and equation 2 represents the separable Gaussian Blur filter for our purpose. The separable Gaussian Blur filtering requires less computation costs. Fig. 3 shows the step images which are created in each pass.

$$F(x,y) = c^2 \sum_i \sum_j exp(-\frac{(x_i - x_a^2)}{2\sigma^2}) exp(-\frac{(y_i - y_a^2)}{2\sigma^2}) \cdot t(x_i, y_j) \tag{1}$$

$$\begin{aligned} I(x,y) &= c \sum_i exp(-\frac{(x_i - x_a^2)}{2\sigma^2}) \cdot t(x_i, y_j) \\ F(x,y) &= c \sum_j exp(-\frac{(y_i - y_a^2)}{2\sigma^2}) \cdot I(x_i, y_j) \end{aligned}, \tag{2}$$

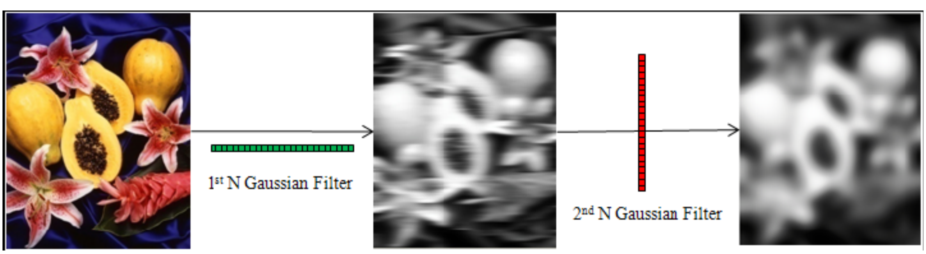


Fig. 3. Images generated by separable Gaussian Blur filter in each pass

3.3 Implementation Using CUDA

We convert the image illustration algorithm by single thread using CPU to multiple threads using current GPU which is capable of parallel computation. The texture array provided by NVIDIA's CUDA is used in order to assign data at each GPU processor and to avoid data bottle neck generated in the SIMD command execution of CUDA. An image is inputted into the texture array and an intermediate value is stored in the illustration process. Also, our implementation of CUDA employs the separable Gaussian Blur filter to reduce computation costs. Below CUDA pseudo codes in Fig. 4 represents these explanations in detail.

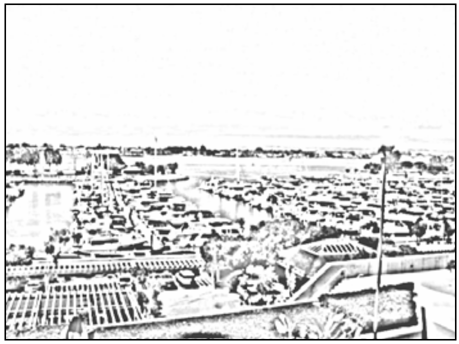
```
Program cudaComputeIllust (*pSrcData, *pDestData, nWidth, nHeight,
fAlphaScale, fThresLumi)
{
   float    *pV1, *pV2, *pV3, *pB;
   cudaArray  *arrTex1, *arrTex2, *arrGuassWeight;
   dim3     threadBlock(blockSize, blockSize, 1);
   dim3     blockGrid(iDivUp(nWidth, threadBlock.x), iDivUp(nHeight,
threadBlock.y), 1);
    // Copy the data in main memory to GPU memory
    cudaMemcpy (pV1, pSrcData, nBytes, cudaMemcpyDeviceToDevice) );
    for (s = 1 ; s<= N ; s++) {
        MakeGussianKernel(s);
        // Horizontal Gaussian Blur filter
        cudaThreadGaussRow2D <<<blockGrid, threadBlock, 0>>> (pV3,
nWidth, nWidth, nHeight, nAnchorX);
        // Vertical  Gaussian Blur filter
        cudaThreadGaussColumn2D <<<blockGrid, threadBlock, 0>>> (pV2,
nWidth, nWidth, nHeight, nAnchorX);
        // Get the differences between orignal & blurred image
        cudaThreadComputeB <<<nHeight, 128 >>> (pB, nWidth, nWidth,
nHeight, fCoeff);
        // Swap original image and blurred image
        swap (pV1, pV2);
    }
   // Get the Final Image
  cudaThreadComputeFinal <<<nHeight, 128 >>> ((Pixel4 *)pDestData,
nWidth, nWidth, nHeight, fThresLumi);
}
```

Fig. 4. CUDA pseudo code for our illustration algorithm

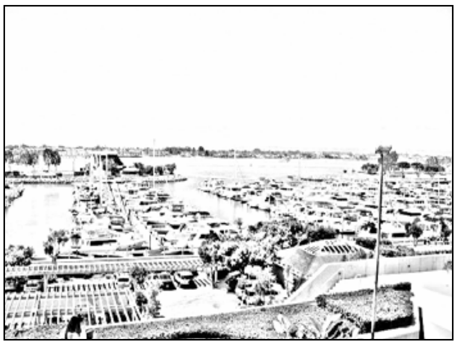
4 Experiments and Results

To find advantages of our real-time video illustration, we conduct two kinds of experiments. The first experiment is about the quality of video illustration. To check the illustration quality, comparison between result images using illustration module in Adobe Photoshop and our results. In Fig. 5, (a) is a source image, (b) is the result image of illustration by Adobe Photoshop, and (c) is the result of our illustration. Comparing the quality of (b) and (c), we can't find much difference. Therefore, when we apply illustration effects to video streams, we can obtain high image quality.

(a) Source Image



(b) Illustration by Adobe Photoshop



(c) Our Illustration Image

Fig. 5. Images generated by separable Gaussian Blur filter in each pass

The second experiment is about real-time performances. In this experiment, we implemented a variety of cases: CPU-based method using OpenCV library [5] and GPU-based method using NVIDIA CUDA. The former is divided into 2 methods according to the form of the Gaussian Blur filter. In short, there are CPU-based method using OpenCV (CPU1), and CPU-based method with the separable Gaussian Blur filter (CPU2), and NVIDIA CUDA method (CUDA). Therefore, we can compare the results of above three methods.

Because performance of an image or a video transformation relies on its image size (resolution), we can test the performances of three methods, using various image sizes. Table 1 represents the number of illustration frames in a video to be generated per second. The image sizes of videos used in our experiments are 320x180, 640x360, 960x540, and 1280x720. All methods were tested on an Intel Q9550 CPU PC with Microsoft Windows XP and a NVIDIA GeForce GTX260.

As shown in Fig. 6, the GPU-based method using NVIDIA CUDA shows more impressive results when comparing with the other methods. Particularly, if the image size of a video is enlarged, the performance difference grows because computation costs of the Gaussian Blur filter increase.

Table 1. The performance evaluation of 3 methods under various video resolutions

Resolution	CPU1	CPU2	CUDA
320 x180	32.26	66.67	148
640x360	8.54	16.13	62.5
960x540	4	7.09	32.36
1280x720	2.56	3.77	21.28

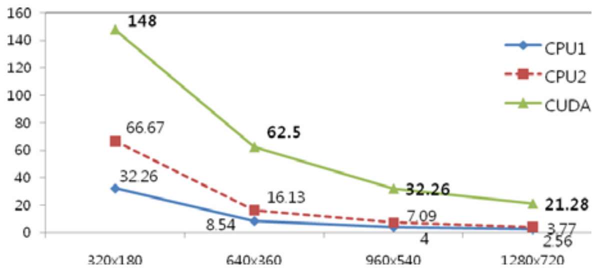


Fig. 6. Comparison of 3 methods (FPS)

5 Conclusion

In this paper, we suggest a video transformation system which is designed to apply illustration effects to video streams directly. It is also implemented by using the GPU hardware with NVIDIA's CUDA for real-time performances. Our system has more effective performances than systems simply with CPU. Excellent results come out in videos of the HD resolution (1280x720) as well. Illustration effects are not influenced in color distributions or in contents of videos because calculation is performed pixel by pixel. Therefore, performances according to the resolutions of videos can be expected for the real application.

In the future, we hope to create real-time video transformation systems with other non-photorealistic rendering effects like cartoon, photo-mosaics, and watercolor.

Acknowledgments. This work was supported by the IT R&D program of MCST/MKE/IITA. [*2008-F-030-02*, Development of Full 3D Reconstruction Technology for Broadcasting Communication Fusion].

References

1. Gooch, B., Erinhard, E., Gooch, A.: Human Facial Illustrations: Creation and Psychophysical Evaluation. ACM Transactions on Graphics 23(1), 27–44 (2004)
2. Winnemöller, H., Olsen, S.C., Gooch, B.: Real-Time Video Abstraction. In: Proceedings of ACM SIGGRAPH 2006, pp. 1221–1226 (2006)
3. Klein, A.W., Grant, T., Finkelstein, A., Cohen, M.F.: Video mosaics. In: Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering, pp. 21–28 (2002)
4. CUDA: Compute Unified Device Architecture, `http://www.nvidia.com/object/cuda-home.html`
5. OpenCV: Open Computer Vision Library, `http://sourceforge.net/projects/opencvlibrary/`

A Distributed Render Farm System for Animation Production

Jiali Yao, Zhigeng Pan*, and Hongxin Zhang

State Key Lab of CAD&CG, Zhejiang University, Hangzhou, 310058, China
{yaojiali,zgpan,zhx}@cad.zju.edu.cn

Abstract. Render farm is widely used in movie industry to solve the long rendering time problem. By parallel computing rendering jobs, render farm can speedup the rendering process in a scalable way. In this paper, we present an efficient design of render farm system name DRFarm in distributed environment. The most important feature of the system is the capacity aware task scheduling strategy. We first introduce the hierarchy tasks subdivision method which ensures flexible merging and dividing tasks. By carefully grouping tasks and dynamically assign them in different modes, the overall parallel rendering time can be reduced by exploiting coherence comparing to conventional methods. Furthermore, to adopt various rendering jobs from different locations, we design a general rendering service interface with unified job definition.

Keywords: parallel rendering, distributed rendering, render farm.

1 Introduction

In movie industry, rendering a full-length animation film cost numerous CPU time. Render farm is built to reduce the rendering time by parallel computing individual frame in distributed environment.

Typical render farm is constructed for animation film rendering. The first full-length animation film, Toy Story, used 117 Sun workstations and the Pixar RenderMan system. The film, Shrek 3, is rendered by more than 4000 HP workstations, where every second requires 3000 hours CPU time. Render farm is also widely used in architecture design, advertising, and the visual effects industry.

Different from other massive parallel systems, render farm emphasizes on cost-effective ratio, and usually adopts contemporary commodity workstations. Although multiprocessor platform performs excellent for parallel rendering, especially interactive ray tracing applications [1], render farm focus on high quality off-line batch rendering.

Our render farm system which named DRFarm is built on commodity multi-core PCs connected by commodity Ethernet infrastructure. In our implementation, we focus on scalability and rendering client utilization by adopting hybrid task schedule method and flexible task grouping strategy.

* Corresponding author.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 264–269, 2009.

2 Related Works

The "render farm" concept has long been applied in movie industry to achieve high quality images by parallel computing. The "Kilauea" renderer parallel computes global illumination algorithms on cost-efficient PCs running Linux, but not considered for real-time purpose. Recently, interactive ray tracing system has been implemented from SGI workstations to even commodity PC clusters [2]. These systems use highly custom optimized renderers. Specially optimized renderers help exploit hardware feature, such as SIMD units [3], but loose certain flexibility.

Commercial render farm management systems include Qube from PipelineFX, Enfuzion from Axceleon, Muster from Virtual Vertex, and Deadline from Frantic Film. There are also open source render farm systems such as drqueue. These systems generally support as much renderers as possible for different users, and focus on service quality. However, such system usually split tasks into single frame, further optimization have to be manually configured if possible.

Online render services are provided by remote render farms, such as RenderRocket, which helps make global utilize of computing resource. Grid [4] or volunteer computing [5] are also suitable for remote rendering. Implementing such remote parallel systems usually requires design network interface and handle data access in complex environment [6].

In our work, we offered novel strategy to handle task schedule. Hierarchy task subdivision definition is introduced to ensure both fine and coarse granularity. With flexible task merging and subdividing method, tasks can be grouped together aware of client's computing capacity to exploit temporal coherence between frames.

3 System Design

3.1 System Overview

Figure 1 is the architecture of the render farm system (DRFarm). In the system, the server manages all the storage and rendering resources, and provides services to remote or local users by general render service (GRS) interface. In the render farm, hardware is connected by local network infrastructures.

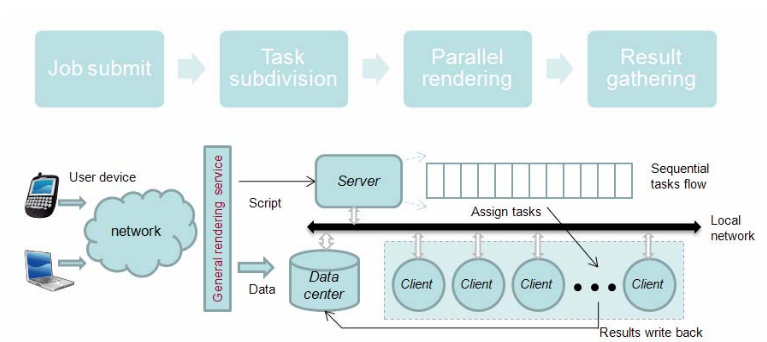


Fig. 1. The architecture of the render farm system

We implement network architecture in C/S mode, and TCP/IP is employed to handle communication. Most other render farm systems use local network file sharing strategy for scene data access, which is based on SMB in windows operation system or Samba in Linux OS. We implement both to avoid network bottleneck.

3.2 Hierarchy Task Subdivision

Renderers used in render farm system are generally based on ray-tracing algorithms which are famous for embarrassing parallelism. When each client has necessary data resources, pixels can be computed independently. For the average user, some popular commercial animation packages employ coarse-grain parallelism to allow rendering of individual frames of an animation across a network of machines. Other renderers such as POV-Ray [8], parallelization scheme works on single images only.

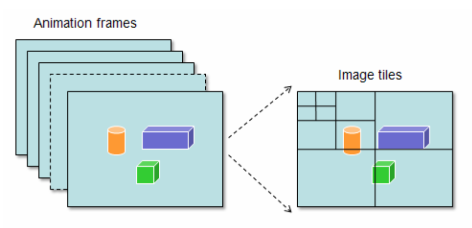


Fig. 2. Temporal and spatial parallelism in animation

Among all the parallel methods, we focus on temporal parallelism between frame sequences and spatial parallelism inside single frame. Since animation frame has inherent corresponding to each other, we employ a hybrid method to dynamically subdivide single frame into finer granular tasks.

Modern hardware architecture benefits from hierarchical arrangement of different levels parallelism. For example, the Cell cluster has three levels hardware parallelisms to exploit [7]. Furthermore, tasks can be expanded and merged in a flexible manner, which is a requirement of capacity aware tasks assign algorithm.

3.3 Dynamic Load Balance

Granularity of tasks greatly affects load balance performance. Conventional method in commercial render farm management system is setting up a task pool and dividing the tasks into a modest but fixed granularity such as one frame or two. With dynamic task subdivision method, we employ capacity aware strategy which will be introduced in experiment section.

We design load balance algorithms for both job completion oriented and client utilization ratio oriented requirements, named active mode and passive mode separately.

In active mode, the server groups and distributes tasks to rendering clients by their rendering capacity. If certain client finished all the tasks assigned, it will search the queue to find unfinished tasks assigned to other clients. In passive mode, all tasks will be pushed into the task queue. Rendering client fetches tasks each time when its current state is idle. The passive mode doesn’t guarantee job completion in coarse granularity.

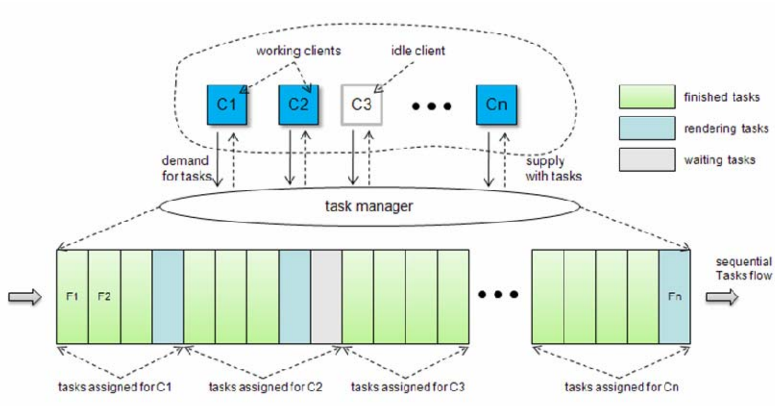


Fig. 3. Dynamic task schedule in active mode

3.4 Unified Job Submission

Supporting various kinds of renderers is the key requirement in the render farm system. We define an abstract level between actual renderer and user, which named general rendering service interface (GRS). After analyzing, we define jobs in quintuple form: Job = {Type, Job decomposition, Renderer, Data access, Result composition}. The five elements are the minimal elements of submitted job script.

Figure 1 illustrates the functionality of GRS. Users located in different area can submit various types of Jobs via GRS interface. A script of the current job will be generated including all necessary items for data access and task subdivision. The script is coded in XML format, thus can be easily decode and transferred into other render farm platform script

4 Experiment

In practice, the computing capacity of rendering clients might be different. Based on the fact that frequent memory operation and communication between rendering clients will cause the system inefficient, we present a capacity aware grouping strategy. Instead of using third party tools to measure performance, we firstly run fine grain tasks to collect overall rendering information of each rendering client. With the heuristic statistics, new tasks are grouped together as coarse grain ones to exploit coherence and reduce loading time.

To illustrate how the flexible tasks subdivision and grouping algorithm work, we task a 48 frames animation (plane fly animation as shown above) for example, which is rendered by Maya renderer. The experiment is running on 6 different workstations.

In this example, compared with scene loading time, rendering time of one frame is very short. The naïve strategy split tasks into single frames thus caused great loading time consumption. The green line is the fixed task group in size of 2. As shown above, capacity aware grouping strategy has flexible task size according to rendering clients’ capacity and outperforms other two conventional strategies.

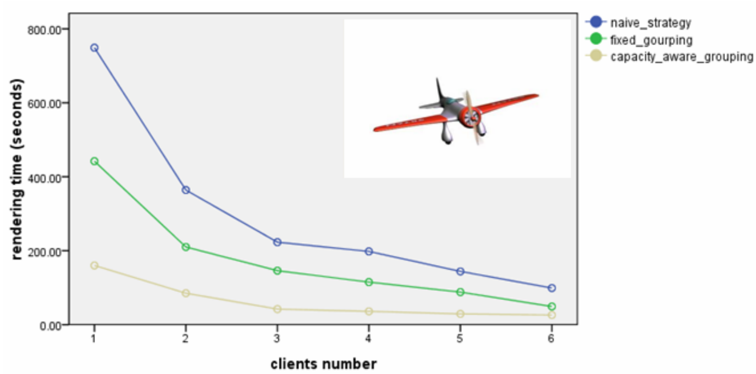


Fig. 4. Rendering time of a 48 frame simple Maya animation

After repeating the naïve while the finest subdivision algorithm several times, we can find that client with more computing power tends to compute more frames. Figure 5 is the average frames rendered at each client.

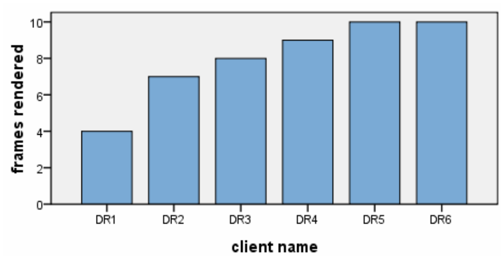


Fig. 5. Client capacity represented by frames computed

Client capacity can be evaluated by clients' rendering history statistics. With client rendering capacity in mind, the system therefore subdivides frames into finest granularity and then merges them by capacity value. As shown in Figure 4, capacity aware strategy can achieve very excellent speed-up.

5 Conclusion and Future Work

In this paper we present the fundamental concept for building a render farm system. We have shown that commodity PC cluster also performances well as a distributed parallel rendering environment. Although increasing the client number in straight forward way will not effectively promote speedup limited by network I/O bandwidth, acceptable scalability can still be achieved using methods like grouping and coherence exploiting.

Currently, render farm is still limited in local area network. However grid and volunteer computing systems can help utilizing remote resources. Future work can focus on idle PCs based remote rendering.

References

1. Bigler, J., Stephens, A., Parker, S.G.: Design for parallel interactive ray tracing systems. In: Proceedings of the IEEE Symposium on Interactive Ray Tracing, pp. 187–195 (2006)
2. Wald, I., Benthin, C., Dietrich, A., Slusallek, P.: Interactive Distributed Ray Tracing on Commodity PC Clusters - State of the Art and Practical Applications. In: Kosch, H., Böszörményi, L., Hellwagner, H. (eds.) Euro-Par 2003. LNCS, vol. 2790, pp. 499–508. Springer, Heidelberg (2003)
3. Benthin, C., Wald, I., Scherbaum, M., Friedrich, H.: Ray tracing on the CELL processor. In: Proceedings of the IEEE Symposium on Interactive Ray Tracing, pp. 15–23 (2006)
4. Foster, I., Kesselman, C.: The Grid 2: Blueprint for a new computing infrastructure. Morgan Kaufmann, San Fransisco (2004)
5. Anderson, D.: BOINC: A System for Public-Resource Computing and Storage grid. In: Fifth IEEE/ACM International Workshop on Grid Computing, pp. 4–10. IEEE Computer Society, Washington (2004)
6. Pan, Z., Shi, J., Zhang, M.: Distributed graphics support for virtual environments. Computers & Graphics 20(2), 191–197 (1996)
7. Komatsu, K., Takizawa, H., Kobayashi, H.: Hierarchical Parallel Processing of Ray Tracing on a Cell Cluster. In: 1st International Workshop on Super Visualization, CD-ROM (2008)
8. POV-Ray documentation, `http://www.povray.org/documentation/view/3.6.0/7/`

Extending the STRADA Framework to Design an AI for ORTS

Laurent Navarro and Vincent Corruble

Laboratoire d'Informatique de Paris 6
Université Pierre et Marie Curie (Paris 6) – CNRS
4, Place Jussieu
75252 Paris Cedex 05
{Laurent.Navarro,Vincent.Corruble}@lip6.fr

Abstract. Strategy games constitute a significant challenge for game AI, as they involve a large number of states, agents and actions. This makes indeed the decision and learning algorithms difficult to design and implement. Many commercial strategy games use scripts in order to simulate intelligence, combined with knowledge which is in principle not accessible to human players, such as the position of the enemy base or the offensive power of its army. Nevertheless, recent research on adaptive techniques has shown promising results. The goal of this paper is to present the extension such a research methodology, named STRADA, so that it is made applicable to the real-time strategy platform ORTS. The adaptations necessary to make STRADA applicable to ORTS are detailed and involve the use of dynamic tactical points and specific training scenario for the learning AI. Two sets of experiments are conducted to evaluate the performances of the new method.

Keywords: Game AI, learning, real-time strategy games.

1 Introduction

The quality of a commercial video game depends largely on its capacity to entertain human players. After having invested significant efforts to increase the graphic quality of their games, making them more realistic, game designers try to focus on improving the gameplay of their products. Nevertheless, the Artificial Intelligence (AI) available in games remains usually limited and predictable, often forcing the players to compete against other humans instead of synthetic entities [3,5].

Some of the most highly used AI techniques in video games, such as Finite State Machines (FSM) or Scripting languages which are powerful solutions, easy to implement, let programmers describe behaviors in a static and somewhat detailed manner. They can lead to realistic behaviors [8], but they are also plagued by complexity [4]. Moreover, their determinism makes them predictable by the human player after a certain amount of runs [5]. A promising evolution to go beyond this limitation is to look for adaptive techniques, where the knowledge necessary to the behavior is not produced by the programmer, but is learned automatically through experience (i.e.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 270–275, 2009.

through playing). It has been proposed for example with Dynamic Scripting [1] which uses weighted rules to adapt scripts. Though some promising results have been obtained in research labs developing learning techniques for games [7, 9], they remain so far underused in commercial games [5].

In the following section, this paper briefly introduces the STRADA framework for an adaptive game AI on which this paper is based, and the ORTS platform which is used as an environment for our experimentation. It then tackles the challenge of adapting the STRADA framework to the ORTS platform for real-time strategy (RTS) games, in particular looking at the question of map analysis and tactical points. The resulting platform is then tested against the winner of the 2007 ORTS competition, and against a random AI. Encouraging results are further improved by proposing the notion of specific training scenarios where the learning AI is set in an environment that favors the acquisition of key game concepts.

2 Background

The goal of the STRADA framework [2] was to propose a generic model for the automatic generation of adaptive strategic behaviors in strategy games. It combines recent AI techniques, like reinforcement learning, with new ideas to handle the large complexity of modern games. Three main axes were explored: a decision-making system architecture based on a military hierarchy and a map analysis algorithm, whose goal were to reduce the complexity of the state and action spaces, and specific combined reinforcement signals which dispatch the information through the hierarchy and help the coordination between the different learning agents.

This approach has been applied to the turn-based game Battleground: Napoleon in Russia (Talonsoft). Experimental results showed that the STRADA approach reaches higher performances than those obtained by the original game AI, and is able to compete with a trained human player. A coherent and adapted military strategy was learned for the two scenarios studied. Only Battleground was used so far for the evaluation of STRADA, while the ambition behind it was to have a somewhat generic framework for strategy games adaptive AI. Trying to test and adapt the STRADA framework on a modern real-time strategy game is an important motivation behind the work presented here.

The ORTS platform is an Open Source project aiming at providing the scientific community with a shared framework for RTS AI testing [6]. It is based on a client–server architecture where all the central data, like the position of all the playing units, are handled by the server. This environment is nearly comparable to commercial real-time strategy games like Command & Conquer (Westwood Studios), but some important differences with *Battleground* (used in past experiments with STRADA) need to be highlighted, as they require specific adaptations explained further in the following section. The random generation of map in ORTS (for each new game) highlights the need for a new approach to the map analysis method proposed in STRADA. The presence of an economy in ORTS, symbolized by the management of resources in order to develop a base and an offensive army, where Battleground focused on the tactical aspect of the conflicts, requires that the new platform adapts the notion of hierarchy previously used in STRADA. Finally, considering that ORTS is a real-time game and that Battleground is turn-based only, the new model has to adapt the learning algorithms to this new complex testing environment.

3 Extension of the STRADA Model

3.1 Hierarchical Structure

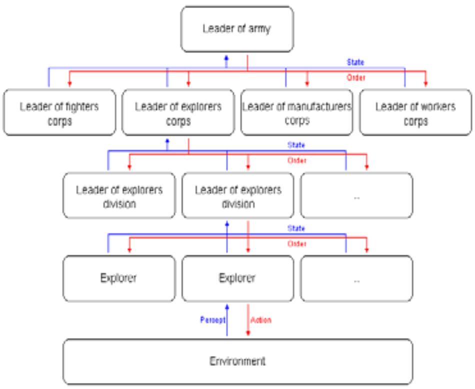


Fig. 1. Military hierarchy from STRADA adapted to ORTS corporations

The platform described here replicates the vertical dimension of the military hierarchy already used in STRADA, from army leader at the top to individual units at the bottom.

However, the presence of *fog of war* in ORTS requires creating special unit dedicated to exploration. Similarly, the simple economy in ORTS requires the implementation of workers for collecting resources and building production structures, and of manufacturers, for creating new units. All those corporations have specific orders and perceptions. Therefore, the new platform introduces a horizontal dimension to the original vertical hierarchy introduced by STRADA with four components (fighters, explorers, manufacturers, and workers) as shown in Figure 1.

3.2 Dynamic Tactical Points

Tactical points are structures created by the platform in real-time to abstract the knowledge acquired while playing. At each round, the engine locates and identifies special areas of interest (see Fig. 2 for an example), mainly defined by the presence of groups of buildings, and extract specific information which is stored to create the memory of the game. Those parameters, which are discretized and normalized, describe (1) the strength of friendly forces, of (2) enemy forces, (3) a risk factor, (4) a force ratio, (5) resource availability.

Thus, tactical points are a combination of those parameters, calculated and modified in real-time by the engine. In this study, the different values used for each setting allows the creation of 36 singular tactical points. Each of them is then combined with the different orders usable by the 3 operating level to create the action space. For the purpose of this study, only buildings can create a tactical point, even if surrounding units offensive and defensive power are represented in its description. Creating tactical points with only troops or landscape singularities would be feasible but adding parameters to their description will exponentially increase the number of allowed tactical points, as well as the size of the action space.

Fig. 2. Example of inference of 6 tactical points during a game

3.3 Reinforcement Signals

The structure of the reinforcement signals used in the new platform is a consequence of the military hierarchy described previously. Most of them are similar to the one illustrated in the STRADA model: the global reward, calculated from the score obtained by the agent mainly by collecting resources and killing opponents, the local reward, specified for each leader of the different corporations, and the order reward, representing how a leader follows the order given by his direct hierarchy.

However, to represent the horizontal axis added by our new framework, a specific local combined reward has been introduced, whose definition is shown below. Its goal is to symbolize the interaction between the different corporations. Finally, the complete reward, named combined reward, is a linear combination of the main rewards explained above:

$$R_{Local\ Combined} = \alpha R_{Fighters} + \beta R_{Explorers} + \gamma R_{Manufacturers} + \delta R_{Workers}$$
$$R_{Combined} = \alpha R_{Global} + \beta R_{Order} + \gamma R_{Combined\ Local}$$

This final reward is used within a SARSA-λ learning algorithm [10]. The combined rewards associated to the state/action couples are memorized using neural networks. Finally, the action selection strategy is based on a Boltzmann-Gibbs probabilistic distribution.

4 Experiments and Results

During this first experiment, the new platform has been trained on the third scenario of the ORTS AIIDE competition, during 5.000 steps. It is opposed both to the *RTSComp07* game AI, winner of the AIIDE07 challenge, and to a *test AI* using the same engine as the new platform but performing a *random* action at each decision cycle. Every 100 runs, an evaluation was performed during 20 games to measure different performance indicators such as the score, the offensive and defensive power of the army, the size of the explored map and the amount of farmed resources.

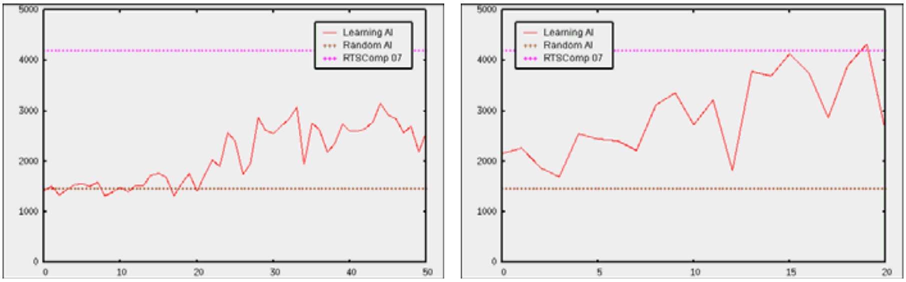


Fig. 3. Evolution of the score function with the number of evaluation steps (a) (left) without using specific training scenarios (b) (right) with specific training scenarios

The score evolution detailed in Figure 3a (left) shows that the platform is able to increase its performance through learning. After 5.000 runs, its score is 78% higher than the one obtained by the *random AI* but 45% lower than the estimated score of *RTSComp07*. Moreover, the platform takes more than 1500 steps before it begins increasing its performance.

After learning, the AI is able to farm resources, explore the map and optimize its global score by creating a few offensive units. Nevertheless, it does not learn to colonize unused resources spots and cannot launch significant assaults against the opponent. These somewhat poor results can be explained by the large amount of stages needed to be completed before being able to develop a massive army. At the opposite, the *RTSComp07* AI rushes the opponent base early in the game.

Following the half-satisfying results described above, an attempt was made to increase the performance of the platform by creating specific training scenarios, which are designed to let the agent acquire important skills without fearing an early attack. Three training scenarios have been designed, each one to learn a specific ability: (1) exploration with an empty map, (2) tactical coordination with an already created assault force and an identified enemy base to target, and (3) army development with a pre-built complete base. Finally, the skilled agent having learned through these three training scenarios has been tested again against *RTSComp07* during 2.000 runs with the same evaluation method as the one defined previously.

This time, the score evolution displayed in Fig. 3b (right) shows an important enhancement of the performance, which is after learning now 235% higher than the one obtained by the *random* AI, and only 16% lower than the estimated score of *RTSComp07*. With the use of the learning scenarios, the agent is now fully able to explore the map, farm resources, manage the production of units regarding the amount of resources collected and produce a massive army. Nevertheless, it is still not able to colonize unused resources spots and efficiently coordinate its army to attack the enemy bases or defend its own base. Let us note that the AI performance is highly dependent on the quality of the scenarios it has been trained on. Those should be further improved to focus on the defaults observed previously.

5 Conclusion

The results analyzed previously showed a great capacity of the new framework to learn a strategy and to increase its performance in a complex STR environment. Exploration and basic economical behaviors have been successfully learned during both experimentations. The use of training scenarios has improved military tactics with the production of a massive infantry army and the emergence of artillery. Nevertheless, no decisive assault is launched on the enemy base and only very few victories against *RTSComp07* have been registered.

As shown previously, those points could be improved by slightly modifying the model and optimizing most of the low-level AI algorithms used in the engine. Moreover, the training scenarios used to increase the agent's initial knowledge appeared to be difficult to design and not completely satisfactory. Another idea would be to replace them by a military doctrine. It would represent what servicemen learned at school as a basis before learning "in battle".

Finally, the performances obtained by this new platform are very encouraging. Most of its aspects can be improved to allow it to produce a high-quality strategy and to be able to defeat most of the script-based AI with its adaptive capacities. Such progress would allow it to be a great opponent in the next ORTS competitions, and ultimately a more entertaining opponent to human players.

References

1. Spronck, P., Ponsen, M., Sprinkhuizen-Kuyper, I., Postma, E.: Adaptive game AI with dynamic scripting. Mach. Learn. 63(3), 217–248 (2006)
2. Madeira, C., Corruble, V.: STRADA: une approche adaptative pour les jeux de stratégie modernes. Revue d'Intelligence Artificielle. Hermès, Lavoisier 23(2-3), 293–326 (2009)
3. Forbus, K.D., Laird, J.: AI and the Entertainment Industry. IEEE Intelligent Systems 17(4), 15–16 (2002)
4. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall, Englewood Cliffs (2002)
5. Nareyek, A.: Artificial Intelligence in Computer Game - State of the Art and Future Directions. ACM Queue 10, 58–65 (2004)
6. Buro, M.: Call for AI Research in RTS Games. In: Proceedings of the Challenges in Game Aritificial Intelligence workshop, AAAI 2004, pp. 139–141 (2004)
7. Marthi, B., Russell, S.J., Latham, D., Guestrin, C.: Concurrent hierarchical reinforcement learning. In: Proceedings of IJCAI 2005, pp. 779–785 (2005)
8. Khoo, A., Hunicke, R., Dunham, G., Trienens, N., Van, M.: FlexBot, Groo, Patton and Hamlet: research using computer games as a platform. In: Eighteenth National Conference on Artificial intelligence, pp. 1002–1002 (2002)
9. Tesauro, G.: Programming backgammon using self-teaching neural nets. Artif. Intell. 134(1-2), 181–199 (2002)
10. Sutton, R., Barto, A.: Reinforcement Learning, An Introduction. MIT Press, Cambridge (1998)

Services in Game Worlds: A Semantic Approach to Improve Object Interaction

Jassin Kessing, Tim Tutenel, and Rafael Bidarra

Computer Graphics Group
Delft University of Technology, The Netherlands
JassinKessing@gmail.com, T.Tutenel@tudelft.nl, R.Bidarra@ewi.tudelft.nl

Abstract. To increase a player's immersion in the game world, its objects should behave as one would reasonably expect. For this, it is now becoming increasingly clear that what game objects really miss is richer semantics, not eye-catching visuals. Current games' lack of semantics is mostly due to the difficulty of game designers to realize such complex objects. This paper proposes a solution to this problem in the form of services, characterizing classes of game objects. An example of this is the service of a vending machine, which exchanges a coin for a soda. A three-phased methodology is presented to incrementally specify and add services to game objects. This approach has been implemented and validated by means of a prototype system, which enables a simple and intuitive definition of services in an integrated environment. It is concluded that game objects aware of their services facilitate more and better object interaction, therefore improving gameplay as well.

Keywords: game worlds, services, semantics, object interaction.

1 Introduction

Look around and you will probably see objects scattered all around the place. If the same room would be used as the virtual environment of a game, one would probably want to see the same objects - or at least some objects - because empty environments are unnatural to walk through. Game environments that are filled with objects will therefore help immerse the player into the game world. By using graphics, animations, and physics, virtual objects could appear as players expect. However, that only accounts for their visual aspect, because most objects in games are still useless, being there for decoration purposes only. Only few objects, which are crucial for the game progress, are made functional.

An example is the role-playing game *The Elder Scrolls IV: Oblivion* [1], where objects can be picked up and used on specific locations to trigger an event, like opening a door with a key. Some objects have an effect on the player's avatar, for instance eating bread to increase the health, or wearing armor to increase the level of defense. In the game *Alone in the Dark* [2], several objects, which look useless at first sight, can be combined to create a functional object: a full battery and an empty flashlight will provide a light in the dark when combined.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 276–281, 2009.

In the examples above, the functionality of each object (including its meaning, roles, etc.) was thought up by the game designer and implemented by the programmer. In the real world, a particular object may assume other functions or roles never anticipated by its designer; with a game object in a virtual world, this is definitely not (yet) the case, and certainly not automatically. This limitation makes it impossible for a player to (make his avatar) interact with a game object in many reasonable ways.

In the fields of linguistics, computer science and psychology, *semantics* is the study of meaning in communication. When focusing on virtual environments for computer games, semantics is the information conveying the meaning of (an object in) a virtual world [3]. A serious problem in current game development is a lack of tools to easily specify and add semantics to objects, resulting in a lack of object semantics in games. With a semantically rich object representation, virtual objects assume behaviors like in the real world, instead of consisting of a geometric model only. This can be illustrated with a few examples. When eaten by a character, an apple will reduce the hunger level of that character; in other words, an apple provides the *service* of satisfying someone's hunger. A coat serves its wearer for warmth. A fire, however, will provide warmth to everyone in the area. And a vending machine has the service to supply cans of soda, but only after it has received money.

The role of semantics in virtual environments is receiving increasing attention [3], but so far not much research has been done on adding semantics to game objects, let alone with the purpose of making them more functional or improving the overall gameplay. This paper focuses on our research efforts to improve the semantics of objects placed within game worlds. In particular, the notion of services is proposed, by which virtual objects get to 'know' about their roles in the world, how they can affect other entities (including the player's avatar or artificial agents), and how others can interact with them. Empowered with the notion of services, objects acquire their own behavior, instead of a purely predefined behavior; they can behave as one expects, and correspondingly one is able to interact with them as one expects, regardless of whether the virtual object is completely imaginary or mimicking some real world object. We believe that this can significantly change and improve the gameplay, as players will be able to express their creativity and find more paths to achieve the same goal. Enabling game developer teams to easily declare services and assign them to object classes, as described in this paper, is a major step towards the ultimate goal of achieving more and better object interaction.

2 Related Work

Smart objects [4] were a successful proposal for adding semantics to virtual objects, dealing with many of the possible user interactions in a virtual environment. Noticeably, smart objects were primarily devised for manipulation, animation, and planning purposes, like grasping, pulling, or rotating (individual parts of) objects. An example is an artificial agent that can open a door by moving its hand to the door knob, using the correct hand posture, and turning the knob. Although smart objects are powerful for these purposes, they lack the information of which services they provide to their users.

Research in artificial intelligence (AI) proposed the notion of ontologies, due to the lack of shareable and reusable knowledge bases. An ontology is an explicit specification of a conceptualization: a representational vocabulary for a shared domain of knowledge, in the form of human-readable and machine-enforceable definitions of classes, relations, functions, and other objects [5]. When placing ontologies in the context of this research, they define the meaning of objects and the relations between them. In ontologies, important relationships are *generalization and inheritance*, where classes are connected, and each subclass inherits the features of its superclass [6]. The class *Car*, for example, has the class *Vehicle* as its parent. Another important relation is *instantiation*, which relates a class with each of the individuals that constitute it. A *Ferrari*, for instance, is a kind of *Car*. In Section 4, the usefulness of these two relations for the creation of objects with services will become apparent.

3 Designing Services

In order to design services for game objects, it can be very useful to analyze how real world objects can be structured and classified. In particular, we can identify the notions of *class* and *attribute*. Each object in the real world can be said to belong to some class, defined as 'a generic description of a collection of entities based on their essential common attributes'. An attribute, in turn, is defined as 'a characteristic of an entity'. Classes, therefore, describe entities, varying from physical objects like 'apples' and 'people', to substances like 'water'. For attributes, one can think of abstract attributes like 'edibility', or physical attributes like 'mass'. Units and states express the values of attributes. The 'mass' attribute could be expressed in the units 'kilograms' or 'ounces', while 'edibility' could be expressed in the states 'edible' or 'inedible'. Units are also required to express substances, because they are not quantifiable in integer values only, unlike physical objects.

The notions introduced above give us a foundation for the definition of services. In the real world, entities have particular functions, and provide services, and this should also be the case for entities in a virtual world; for example, a coat provides the service of supplying warmth, but only when it is worn. We define a *service* as 'the capacity of an entity to perform an action, possibly subject to some requirements'.

An *action* can then be described as 'a process performed by an entity, yielding some attribute value changes or (new) entities'. Actions are best illustrated by some examples. A service of a heater is to heat the entities in the surrounding area. This means that the values of their temperature attribute rise; see Fig. 1. Attribute value changes do not have to occur to target entities only; they can also affect the actor. Consider an avatar punching an enemy, which lowers the enemy's health, but also increases the avatar's fatigue. The service of a vending

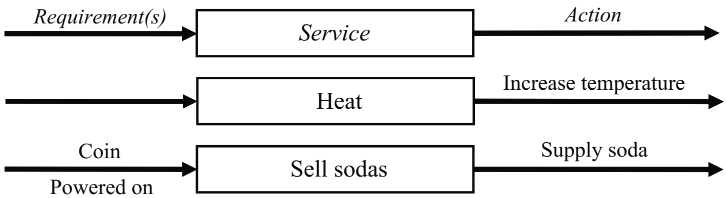


Fig. 1. A generic service, a service of a heater, and a service of a vending machine

machine, supplying a soda, is a good example of an action yielding an entity; see again Fig. 1. This soda is an entity that is supplied from the inventory of the vending machine. However, it is not necessary that the actor always has a stock of existing entities that can be supplied, as an entity's action can also yield new entities. An example is a saw machine that requires trunks, and processes them into wooden planks, which are new entities. This process leads us to the notion of *service requirements*: they can be either actions (e.g. the coat should be worn, and the saw machine should be given trunks) or some attribute constraints, as for example a range of values/states (e.g. electrical devices should be powered on before performing an action, and the fatigue level of the avatar should not be too high before being able to fight).

From the examples above, there are four important elements that should also be taken into account when designing services. First, quantities are essential to indicate how many entities are exchanged during an action, or in which amount an entity is exchanged, in case of substances. Second, temporal properties are relevant, because they indicate the duration of a service, which could be a one-time event, or last for some amount of time. Third, spatial properties indicate who or what is affected by a service, e.g. the consumer, or all entities within a certain radius. Finally, a sequence of interaction steps indicates the order in which requirements should be met before performing an action.

4 Services Put to Work: A Three-Phased Approach

The concepts developed in the previous sections have been implemented in a prototype system which supports the definition of services for game objects step by step. This system covers the three main phases that were identified in the object design process: (i) a *specification phase*, in which generic classes are specified in a library, (ii) a *customization phase*, where a selection of classes from that library is customized into concrete game-specific classes, and (iii) an *instantiation phase*, where object instances of these game-specific classes are placed in a game world. Figure 2 gives an overview of these three phases.

The key idea of the specification phase is to create a library of generic semantic classes. By designing generic classes that can be used in all kinds of virtual worlds, consistency and reusability are stimulated, and thus development time reduced. In this phase, libraries of classes, attributes, units, states, and actions

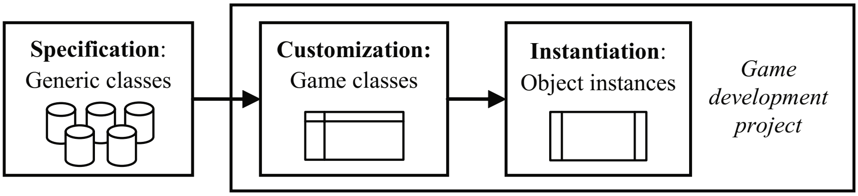


Fig. 2. A phased approach with generic classes in the specification phase, game-specific classes in the customization phase, and object instances in the instantiation phase

are created. Relations can be established among these components, and for each class, services can be defined in order to specify its semantics. By applying inheritance, a class hierarchy is developed, with attributes and services that have been assigned to a class being inherited by all its children. In this way, for example, an attribute like 'mass' does not have to be defined for each single class, but to the 'physical object' class only. To populate these libraries, the WordNet database [7] was used, as it contains many nouns and verbs in the English language, being therefore useful for many possible classes, attributes, actions, etc.

In contrast with the specification phase, the other two phases are not generic. Instead, customization and instantiation play a central role during each particular game project, which typically has its own unique environment, object style and desired behavior, etc. In the customization phase, specific game classes are derived from the generic classes from the first phase, thus automatically inheriting all their generic semantics, including their set of attributes, and also the services they provide. It suffices then to customize the specific behavior desired for this particular project, including their specific attribute values, e.g. quantities and temporal properties. The customization phase is also the right time to assign the project-specific 3D models to the relevant game classes. Finally, in the instantiation phase, instances of the customized game classes can be created and placed inside a game world, which is done by means of a level editor specifically created for this purpose.

5 Conclusions

Despite exuberant visuals, most current games considerably lack proper semantics in the objects populating their virtual worlds. This is partly because designing semantic objects poses especially difficult challenges, including the inherent complexity of maintaining and scaling all interactions among such objects. This paper presented a solution to that problem in the form of services, specified as characteristics of classes of objects. A three-phased methodology has been presented that enables a game development team to incrementally specify and add services to game objects. This approach has been implemented and validated by means of a prototype system, providing an integrated environment which effectively supports a simple and intuitive definition of services. Among the numerous advantages of this approach, among them (i) it promotes reusability of previously specified object semantics, (ii) it easily supports behavior customization as required by each specific game, and (iii) it seamlessly blends with our semantics engine, charged with all service handling during the game (analogously to what a physics engine does with in-game physics).

We believe that enabling designers to create game objects that are aware of each other's services will be instrumental to achieve more and better object interaction. This in turn is considered one of the key conditions to significantly improve gameplay. However, it should also be stressed that object semantics isn't but a (powerful) means to serve the gameplay. In particular, it will never automatically make dispensable the creative work of designers. On the contrary, care should be taken to avoid overloading objects with superfluous semantics, as semantics make virtual objects not only more realistic, but more complex as well, which could end up undermining the gameplay. Therefore, it is the task of the game designer to seek a balance between achieving realism and good gameplay. The approach presented here, giving designers the possibility to include realistic semantics by means of services, while keeping much control on the fine-tuning of the behavior of their objects, is a valuable aid in that direction.

In the future, we would like to experiment with coupling our semantics engine with a game AI system, so that artificial agents can make use of services as well, in addition to players' avatars.

Acknowledgement. This research has been supported by the GATE project, funded by the Netherlands Organization for Scientific Research (NWO) and the Netherlands ICT Research and Innovation Authority (ICT Regie).

References

1. Bethesda Game Studios: The Elder Scrolls IV: Oblivion. Bethesda Softworks (2006)
2. Eden Studios: Alone in the Dark. Atari (2008)
3. Tutenel, T., Bidarra, R., Smelik, R.M., de Kraker, K.J.: The Role of Semantics in Games and Simulations. Computers in Entertainment 6(4) (2008)
4. Kallmann, M., Thalmann, D.: Modeling Objects for Interaction Tasks. In: Proceedings of the 9th Eurographics Workshop on Animation and Simulation, Lisbon, pp. 73–86 (1998)
5. Gruber, T.R.: A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition 5(2), 199–220 (1993)
6. Huhns, M.N., Singh, M.P.: Ontologies for Agents. IEEE Internet Computing 1(6), 81–83 (1997)
7. Miller, G.: WordNet: A Lexical Database for English. Communications of the ACM 38(11), 39–41 (1995)

Glasses-Free 3D Image Viewer by Handmade DIY Craft

Takashi Ohara and Kunio Sakamoto

Department of Intelligence and Informatics, Konan University
8-9-1 Okamoto, Higashinada, Kobe 658-8501, Japan

Abstract. We developed a glasses-free 3D stereoscopic display using an LCD display panel, a view control film and a 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 the view control film can give us glasses-free 3D viewing. As the result, the observer can watch overlapped stereoscopic images for left and right eyes without special glasses such as polarized glasses.

Keywords: 3D imaging, polarized glasses, overlapping stereoscopic images, optical grating film, 3D adapter, stereoscope.

1 View Control for Stereo Viewing

To overlap left and right images, the authors use an optical film "SOLF™" of 3M Company. The SOLF 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. 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 superimposing stereoscopic images by adjusting the interval between an optical sheet and image plane as shown in Fig. 1. 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. If this 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. If you are content to wear glasses, this technology enables us to enjoy 3D image viewing using polarized glasses. In order to improve a bother, it is necessary to separate a side-by-side stereoscopic image into left and right eyes.

To deliver left and right images into appropriate eyes, we use a view control film "LUMISTY™" of Sumitomo Chemical Co., Ltd. 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 of LUMISTY is that it can be either transparent or opaque, so that it looks either like transparent or frosted glass, depending on the angle of sight. It is an adhesive-type transparent plastic film which can be used simply by sticking onto awindowpane, and it does not cut out any of the light coming through the window.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 282–283, 2009.

Fig. 1. Superimposing stereo images

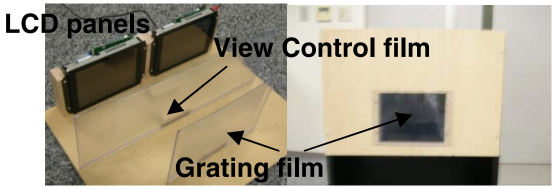


Fig. 2. Appearance of 3D display (KNA-20X)

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 LUMISTY visibility control as shown in Fig. 4, 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 an angle of θ. Let's design the optical layout assuming that 15-inch display panel is used. The width of the 15-inch panel is approximately 280mm. 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 ray with an angle β passes into the right eye through the grating film. If the view control angle θ is α to β ($\alpha < \theta < \beta$), the left image is observerd 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. In case of the 15-inch panel, the angle α is 13.37 deg and the angle β is 25.64 deg. The LUMISTY film has many kinds of characteristics; *e.g.*, opaque from front side, one direction, two directions and so on. The grade MFY-2555 is opaque from one direction when the ray is encountering the film with the angle more than 25 deg. Using this MFY-2555 (θ=+25 deg), 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 view control film 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 stereo viewing. We have developed the prototype glasses-free stereoscopic 3D display using two commercial LCD panels for playing 3D contents by portable DVD players as shown in Fig. 2. In this display, observers can view the 3D images by the binocular stereo viewing without special glasses.

(a) viewing at front

(b) viewing from side

Fig. 3. View control film

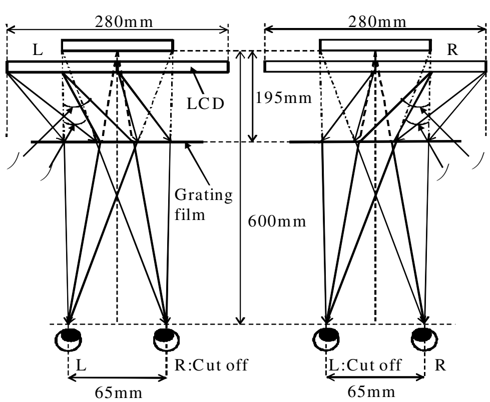


Fig. 4. Optical design for 3D viewing

Monocular 3D Vision Using Real-Time Generated Scene with Depth of Field Effect

Takashi Hosomi and Kunio Sakamoto

Department of Intelligence and Informatics, Konan University
8-9-1 Okamoto, Higashinada, Kobe 658-8501, Japan

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. The authors have developed a monocular 3D vision system with accommodation mechanism, which is useful function for perceiving depth.

Keywords: monocular stereoscopic display, real-time stereogram, 3-D display.

1 Concept

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.

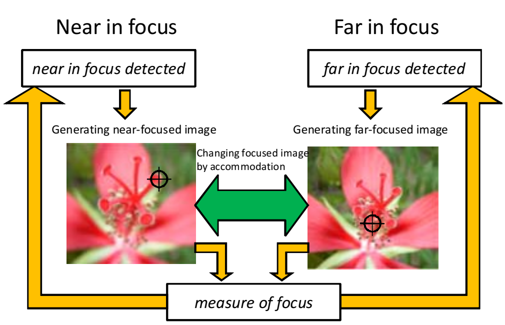


Fig. 1. Flow of generating focused image

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 284–285, 2009.

2 Display System

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.

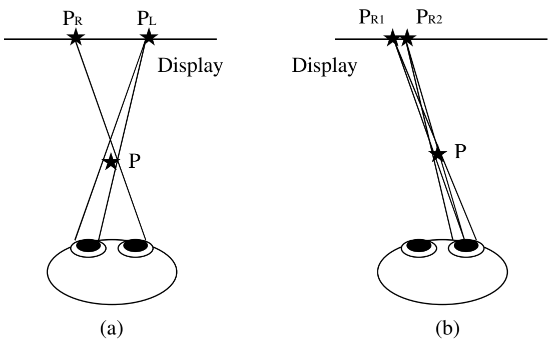


Fig. 2. The principle of a stereogram

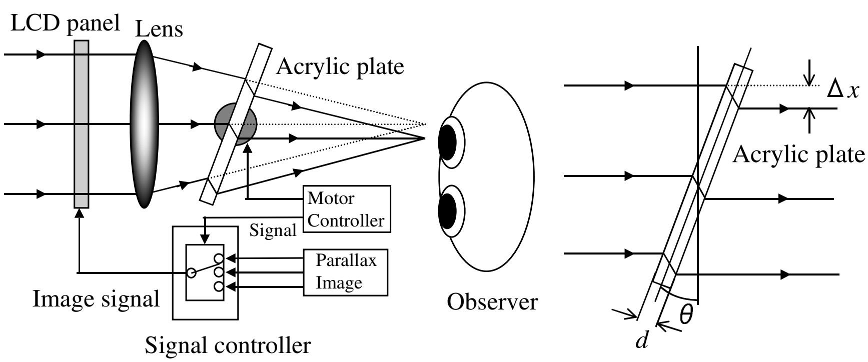


Fig. 3. Optical layout of the monocular 3-D display

RFID Painting Demonstration

Olivier Haberman[1], Romain Pellerin[2], Eric Gressier-Soudan[2], and Ugo Haberman[3]

[1] 17 rue Casimir Perier, 75007 Paris, France
olivier.haberman@orange.fr
[2] CNAM-CEDRIC, 292 rue St Martin, 75 141 Paris Cedex 03, France
{romain.pellerin,eric.gressier_soudan}@cnam.fr
[3] Hippocad, 16 rue du Château, 77300 Fontainebleau, France
ugo.haberman@hippocad.com

Abstract. This demonstration challenges the conventional art experience. It brings together the fields of art, science, and software integration. The goal was to create a new kind of painting based on embedded technology, experimenting with a new media: paintings augmented with RFIDs. The aim was to achieve interaction between the artist's paintings and art gallery visitors.

Keywords: RFID, interactivity, art, painting, mobile phone.

The goal of the demonstration described here is to show how embedded technology could help the artist to create a better interaction between his paintings and art gallery visitors. RFID (Radio Frequency Identification) tags and mobile phone handsets came out very quickly. The artist's requirements were that the painting was free of any visual technology and that the content sent to the visitor could be changed overtime… RFID Tags were an appropriate choice to be the basis of our proposal. RFIDs are easy to use. They can be read or written through an NFC (Near Field Communications) enabled off-the-shelf inexpensive mobile handset (Nokia NFC6131). RFID Tags can be audio and text-rich information providers. They can also carry references to web sites allowing downloads of dynamic content. The communication framework that has been used in the smart painting is part of the uGASP project (gasp.ow2.org). uGASP is a middleware for ubiquitous games. This information can be altered over time by the artist, an additional dynamic and unprecedented capability. RFIDs and NFC mobile phones are a new flexible communication tool that can deliver the living voice and words of the artist thereby closing the gap between artists and the public. The technical part of the application presented here was developed in a joint project with the Hippocad Company, and the CEDRIC research lab of CNAM Paris. O. Haberman is a Painter. As an artist he has observed both at his own and other artist's exhibitions that it would be an advantage to have a direct link between the artist and the public. A way of achieving this for a wide audience with a simple service is to use the latest generation NFC Tag reader mobile phone. It promises the beginning of an enriched interaction with the artist's world.

The RFID based painting described here is provided at the demonstration. Passive tags are inserted directly inside the painting. Seven RFID Tags have been used inside the painting. Their location is indicated by symbols, elements integrated into the work

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 286–287, 2009.

which are easily discernible, regardless of age, culture or language of the viewer. But the RFID tags remain invisible to the public. Approaching a mobile phone from a RFID Tag (less than 2 cm), the phone automatically downloads content, using audio and video programs without the need for user intervention except pushing the "ok" button. The demo shows the various point of interactivity that we have developed to date. The subject of this semi-abstract piece is Communication. The Roman numeral VIII refers to the relationship of men with their universe, the eight planets in our solar system, this allegory is represented by a photograph showing an abstract space environment. The title of the painting: "Infinite Abstraction," which refers to the invisible and infinite dimension of Communication in particular, but also of art in general, inscribed in what evokes the screen of a phone, itself element of communication, enables to send to the viewer a telephone conversation between two children, and addresses the themes of the spoken word among peoples as well as infancy, the future development of our world of Communication. The symbol Ω stamped on the piece, last letter of the Greek alphabet, used here to express the infinite possibilities of correspondence, as opposed to the start of the first communication between men, and whose graphic resonance includes a stylized bridge, invites the viewer to understand the angled triangle of the piece symbolizing a bridge between people, a physical communication link, an architectural, but also a disembodied though very real link via conversation. The number "8", chosen for the gesture of loops that can be reproduced ad infinitum, and which, overturned, brings back to the title (∞: infinite), refers to an abstract photo in which the photographer has chosen to serve the timelessness of communication. The handwritten letter, which occupies the central place in the piece, and historical means of communication, speaks directly to the public to connect with the voice of the author the symbolism used in the creation of this painting. At last, the signature is used to display information on the identity of the piece and of the painter. Thus, this semi-abstract and symbolic piece of art offers a new reading through overt gestures towards and the interaction of many additional media that give a concrete meaning to all the elements that have made the creation of the work. In addition, each of the media has been brought to life by an artist or designer using a different mode of expression. A photographer was invited to give his analysis on his own vision of communication, children have chosen their own grid analysis of the topic and suggested a soundtrack of which they themselves have been the creators and actors, the author was given a space for an audio address to the public to attach permanently to the work the motivations which energized his thought. An additional text in the piece itself includes all the information concerning the author and the identity of the work.

The painting was shown to a range of 48 people between the ages of 12 to 56 years old, of different educational levels and socio-professional categories and with varying levels of interest in art. It allowed us to study the impact of the interaction between a painting, photos, audio messages and text on people at a painting exhibition. Before making the experiment, users have been taught for ten minutes on how the mobile phone has to be used. This approach to art/technology has enabled communication between the artist, other artists and the public.

Development and Evaluation of a Digital Vegetation Interaction Game for Children

Akiko Deguchi[1], Shigenori Inagaki[2], Fusako Kusunoki[3], Etsuji Yamaguchi[4], Yoshiaki Takeda[2], and Masanori Sugimoto[5]

[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, takedayo@kobe-u.ac.jp
[3] Faculty of Art and Design, Tama Art University
kusunoki@tamabi.ac.jp
[4] Faculty of Education and Culture, University of Miyazaki
etuji@cc.miyazaki-u.ac.jp
[5] Graduate School of Engineering, University of Tokyo
sugi@itl.t.u-tokyo.ac.jp

Abstract. In this study, we develop a new digital sugoroku game that portray the phenomenon of vegetation succession in a forest. The results of the experimental evaluation showed that the game was effective in stimulating the interest of the students who participated in the game.

Keywords: Interaction game, Digital sugoroku, Environmental learning, Vegetation succession.

1 Introduction

In the field of science education, because Games can provide fun-fill learning environment, there is plenty of previous research on the subject of introducing digital games to support learning [1] [2]. In this study, we develop a new digital board game to support learning related to the phenomenon of vegetation succession, which can provide fun-fill learning environment regarding the complicated topics in the field of environmental learning.

2 Design and Development of the Vegetation Interaction Game

The vegetation succession game is a digital sugoroku board game that works with Adobe AIR (Adobe Integrated Runtime). We used Mt. Rokko located in the outer reaches of Kobe city in Japan, as a background for this game.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 288–289, 2009.

Figure 1 is the main window of the digital game. 6 pieces represent 6 characteristic plants that grow in the Mt. Rokko region. The surrounding part of the board is the grid area of the sugoroku board. In the central part, there is an event cards area a direction window to move pieces, and a visualization window to show vegetation succession according to the progress of the game. We set the event cards to correspond to the kinds of disturbances.

Fig. 1. Main window of sugoroku board

3 Evaluation of the Vegetation Interaction Game

To evaluate the effectiveness of the digital game in students' learning about the concept of vegetation succession in a joyful way, we conducted an experimental evaluation in a university in Japan. The subjects comprised 18 graduate and undergraduate students who don't major in botany.

A questionnaire survey including 4 items was conducted after the experiment. The subjects were asked to answer the questions using a 4-point scale that ranged from "I think so" to "I don't think so." It took about five minutes for each subject to answer the questions. The results show that for the item "I learned in a joyful way," the positive evaluation significantly outnumbered the negative (p<.01). Further, it was revealed that for the items of understanding of the disturbance, positive evaluations significantly outnumbered the negative ones (p<.01). These results show that this game was effective in supporting the students' interest and learning in a joyful way.

Acknowledgement

This research is supported by Grants-in-Aid for Scientific Research (B) for Masanori Sugimoto.

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., et al.: CarettaKids: A System for Supporting Children's Face-to-Face Collaborative Learning by Integrating Personal and Shared Spaces. In: Proc. of IDC 2006, Tampere, Finland, pp. 45–48 (2006)

4-Views Display System for Collaborative Tasks on Round Table

Mitsuru Okumura and Kunio Sakamoto

Department of Intelligence and Informatics, Konan University
8-9-1 Okamoto, Higashinada, Kobe 658-8501, Japan

Abstract. This paper describes 4-views 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 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 tabletop display system for collaborative tasks cooperated by two users. This tabletop display can provide different images to two users surrounding the system utilizing the image splitting technologies for displaying a stereoscopic 3D image. The viewing zones of this display are generated at both sides in front of observers' eyes. But screens on the monitor cannot be viewed correctly by all users from any direction. Thus, conventional display systems enable users not to do collaborative tasks on the round table. To solve this problem, we developed new viewing systems.

Keywords: all around viewing, group work, grating film, table-top display.

1 All-Around Viewing System

To enable all-around viewing from four directions, a virtual screen is generated and floating above the top level of an actual display panel. Many techniques have been ever proposed in order to float images and locate pseudo images at different places from original positions. To simplify an optical layout, the authors utilize a grating sheet. The grating sheet provides a diffracted image which is arranged under or above 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. Our developed a 4-views display system consists of four LCD panels and a square pyramidal optical screen. Fig. 1 shows a display unit and its layout constituted of four panels. Fig. 2 shows a square pyramid for generating four virtual screens above original positions of the actual display panels. An LCD display produces their image by having a liquid crystal layer that when a current runs through the pixel, it turns on that shade of color. The problem with the liquid crystal is that this color can only be accurately represented when viewed straight on. The further away from a perpendicular viewing angle, the color will tend to wash out. Thus, LCD displays have a limited viewing angle. As an observer watches a viewing screen on the panel with overlooking through the grating sheet, it loses contrast and becomes hard to read at out of the viewing angle as shown in Fig. 3(a). The snapshot of Fig. 3(a) is

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 290–291, 2009.

taken from upward diagonal direction. To correct visual clarity of an LCD's viewing screen, the LCD panels are covered with the grating sheet because a grating diffracts a ray into upwards and downwards. Although the snapshot of Fig. 3(b) is taken from upwards at same angle, it has more contrast and is easier to read at out of the viewing angle using the grating sheet.

A square pyramidal optical screen consists of four grating sheets on its surface so that the four virtual viewing screens are floating in the air. We shaped the viewing screen into a solid pyramid in order to enable all-around viewing from any directions. The grating sheet on the pyramid surface drifts a viewing screen in front of an observer's face. But an observer perceives not only a front view but also neighboring views. To solve this trouble, it is necessary to block lights of both neighboring views. Although louver film can also restrict a viewing angle, we utilized polarizers in consideration for easiness to obtain. The polarizer can switch whether a light goes through or not. If polarizers oriented at 90 degrees to each other, no light gets through. Meanwhile the polarizer passes a light wave when arranged for the same directions. As shown in Fig. 1 and Fig. 2, the LCD panel and the pyramid surface are covered by polarizers. For example, the horizontally oriented polarizers are attached on north and south display panels and east and west panels are covered with the vertical polarizers. North and south pyramid surfaces are also covered with the horizontal polarizers. East and west surfaces have the vertical polarizers too. The lights of both neighboring LCDs are blocked by the polarizers because polarizers on the LCD panel and the pyramid surface are oriented at 90 degrees to each other. An observer can perceive the only forward display through polarizers which have the same direction of polarization.

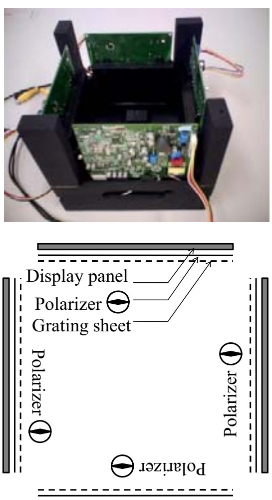


Fig. 1. Display unit

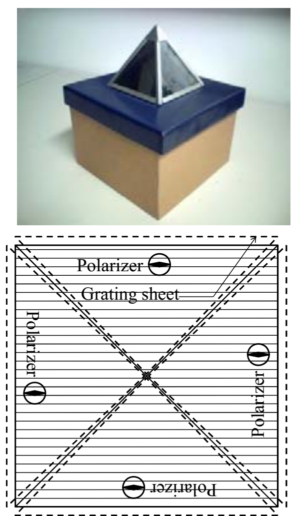


Fig. 2. Square pyramid

(a) no grating sheet

(b) grating attached

Fig. 3. Grating sheet

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) and also by a grant from the Hyogo Science and Technology Association.

Invisible Two-Dimensional Code Display for Additional Information

Tomofumi Yamanari and Kunio Sakamoto

Department of Intelligence and Informatics, Konan University
8-9-1 Okamoto, Higashinada, Kobe 658-8501, Japan

Abstract. The authors have researched a support system of the reminiscence and life review activity. This support system consists of an interactive tabletop display and interface system. Many interaction systems are proposed until now. An invisible code is one of the useful technologies for a computer interaction. The invisible codes provide us with an operating environment using a pen-like device. However, this technology is applied to the only paper media. The authors think we want to realize an interaction using the invisible code on an electrical media. In this paper, we propose a method to display invisible codes using LCD panels and to detect a polarized symbol image with a conventional CCD camera.

Keywords: 2D code, group work, polarized invisible code, polarized light control, table-top display.

1 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 our interaction display system 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.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 292–294, 2009.

A 1/4 wave plate is used to turn-polarized light into circularly polarized light and vice versa as shown in Fig. 3. The final LC layer and this 1/4 wave plate output left or right circular polarized light waves. This difference of rotating direction 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.

(a) view of normal camera (b) thru horizontal polarizer (c) thru vertical polarizer

Fig. 1. Invisible display unit

At the detection, the polarized symbol pattern images are observable through the 1/4 wave plate and the polarizer because the combination of these optical elements blocks the wave or not as shown in Fig. 4. 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.

To simplify optical techniques, Fig. 5 shows an illustration in which a structure involves no 1/4 wave plate. This figure 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. If the polarizer rotates 45 degrees, the detector perceives no information. A 1/4 wave plate can solve this problem. The 1/4 wave plate can cancel this direction dependence. Using this invisible code, the display system provides all users with visual information and assistance like an audio guide if the user needs a support and it can realize the adaptive interface.

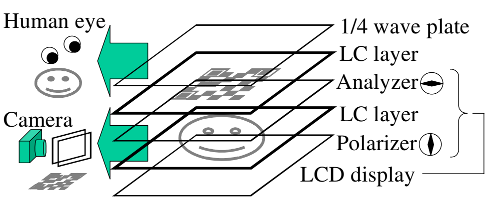


Fig. 2. Principle of generating invisible code

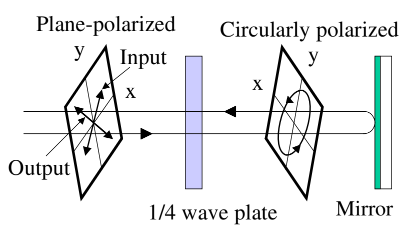


Fig. 3. 1/4 wave plate

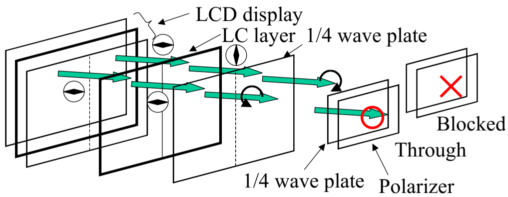


Fig. 4. Principle of hidden code detection

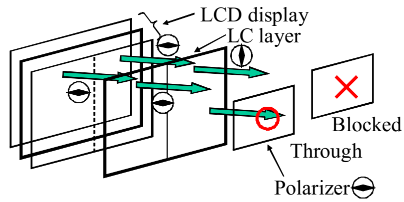


Fig. 5. Basic technology

MobiSpell: Educational Mobile Game Design and Development for Teaching Spelling to Young Children

Menelaos Bakopoulos and Sofia Tsekeridou

Athens Information Technology, 0.8km Markopoulou Ave., 19002 Peania, Athens, Greece
{mbak,sots}@ait.edu.gr

Abstract. A 3D educational mobile game is designed and implemented with the aim to teach hard-to-spell words to young children exploiting their natural affinity towards games entertainment in order to increase the likelihood of useful knowledge acquisition while having fun. The educational game features a fighter plane style environment in which spelling is accomplished by shooting down letters. Incentives and motivation such as points and medals encourage learning and motivate children to develop skills. The J2ME Mobile 3D Graphics library is used with 3D models developed in 3DS MAX, and development taking place in Netbeans.

Keywords: educational game, mobile game, game design, game development.

1 Introduction, Goals, and Design Principles

Rapid technological development has led to modern mobile platforms having equivalent computational power with the early nineties PCs – this has set the grounds to develop entertaining educational mobile games leading it to be a rapidly developing field all over the world.

In this paper, we present an educational mobile game in which a child takes the role of a pilot being tasked with the mission to shoot down various letters and opponents necessary to spell a word correctly. The child is guided by various multimedia cues (such as photos) and blanks representing letters as to which letter should be spelled. In the sequel, we elaborate on key aspects considered in MobiSpell's design referring to deployed technologies and theories, and continue by describing the basic implementation, with conclusions for future work.

MobiSpell's game design has focused on achieving an increased player experience, fun and at the same time learning. Enjoyment directly affects the time spent playing and hence the knowledge retention a child will achieve [1], which made it the primary design principle. Motivation mechanisms such as rewards in the form of praise, points and power-ups have been included as an integral part of the game in order to stimulate game play (Figure 1) with enemies and a timer to play against providing a *competitive element* [2] to increase enjoyment.

An AI component has been included based on a defined state machine interoperating with autonomous moving agents [3] in order to provide an additional competitive element.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 295–296, 2009.

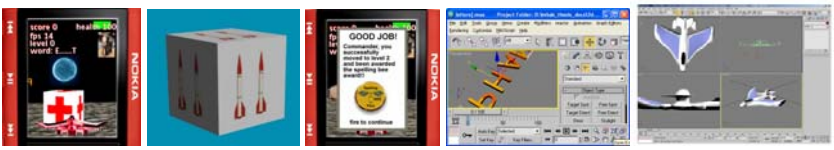


Fig. 1. Health power-up (1), shot power-up (2) , and medal (3) are examples of reward and incentive mechanisms. Letters and 3D models designed in 3DS Max should be pleasing (4,5)

Finally, except for the application of motivational mechanisms, significant effort has been made on providing a pleasant 3D environment since this also plays a role in user satisfaction.

2 Implementation, Performance and Conclusions

For MobiSpell implementation, Netbeans Mobility Pack, 3DS Max, Adobe Firefox and the Hi-Corp M3G Plug-in are utilized. Multimedia content is created in a three step process: 2D textures are first created in Adobe Fireworks, 3D Models are then created using the textures, the models are exported into the J2ME Mobile 3D Graphics format and finally imported in J2ME. Extensive use is made of the M3G library to provide the 3D environment.

The performance of the game was acceptable on a Nokia 6085 cell phone as long as less than two enemy planes and 15 letters existed within the level. Optimization strategies such as using public variables, using a pool design pattern for bullets, and avoiding object oriented encapsulation strategies reduce overhead.

Educational games are effective towards promoting learning - used in conjunction with mobile technology increase all-the-time access to novel learning paradigms. Future work will focus on game extension adding non visual multimedia such as sound (e.g. a spoken word) and optimizing code.

References

1. Sweetser, P., Wyeth, P.: GameFlow: A Model for Evaluating Player Enjoyment in Games. ACM Computers in Entertainment, Article (2005)
2. Williams, R.B., Clippinger, C.A.: Aggression, competition and computer games: Computer and human opponents. Computers in Human Behavior 18, 495–506 (2002)
3. Bakopoulos, M.: A 3D J2ME Game Utilizing Autonomous Moving Agents. In: Tsekeridou, S., Cheok, A.D., Giannakis, K., Karigiannis, J. (eds.) Proceedings of the 3rd international Conference on Digital Interactive Media in Entertainment and Arts, pp. 513–514. ACM, New York (2008)

Live Demonstration of the Pervasive Game “GPS Joker”

Ivo Flammer[1] and David Guyard[2]

[1] XiLabs, Urban Game Studio, 12 rue Vivienne, 75002 Paris, France
ivo.flammer@xilabs.fr
[2] SFR, 1 Place Carpeaux - Tour Séquoia, 92915 Paris La Défense, France
david.guyard@fr.sfr.com

Abstract. From the virtual to the real, everybody is looking for the Joker... your mobile phone guides you through the city. Be the first one to win the Joker. Enter the game and play your session here in Paris!

Keywords: pervasive games, ubiquitous games, urban games.

1 GPS Joker

The mobile telephone operator SFR [1] and the urban game studio XiLabs [2] developed the permanent and persistent pervasive game GPS Joker [3].

Game idea	Find the maximum of “Oblocs” to win the big price at the flashmob.
Game type	Multiplayer LBS Game. Permanent play. Persistent game-world.
Game scenario	

Game alternates between two game phases: a permanent phase and an event-type phase. In the permanent phase you increase your hero’s strength by searching and snatching virtual objects in your city. As you advance into higher game levels, the objects get more and more difficult to catch.

Every two weeks, a flashmob is organized. If you get there on time, you invest your player’s strength in the common pot in order to increase the price payout. If you are the lucky one, you win the big price after the countdown. Otherwise, you might just share a glass with the other players on spot.

References

1. http://www.sfr.com
2. http://www.xilabs.fr
3. http://www.ateliersfr.fr/beta-tests/gps-joker

S. Natkin and J. Dupire (Eds.): ICEC 5709, p. 297, 2009.

Rapid Interactive Installation Development Using Robust Computer Vision and Image-Based Rendering

Denis Perevalov

Institute of Mathematics and Mechanics, Ekaterinburg, Russia
denis.perevalov@mail.ru

Abstract. The paper describes a technique which lets designers implement artistic ideas rapidly into an autonomous interactive art system. The technique consists of two parts: the fixed set of computer vision algorithms and image-based rendering with branching animation sequences.

Despite some limitations, the technique provides easy implementation of the wide range of fascinating interactive scenes.

The availability of low-cost desktops, compact computing systems, digital cameras and TV panels together with existing sophisticated algorithms and software provides an opportunity for spreading the autonomous interactive art installations widely into everyday life. A lot of systems and toolkits are available for developing and deploying such installations [1].

But it is hard to find a system which provides the designer with a fast and efficient way to implement artistic ideas into the interactive form. So there is a demand for a software platform which would allow to implement a wide range of the interactive scenes in a simple way. In this paper, the technique for constructing such a platform is proposed.

Usually an autonomous interactive installation plays some scene while perceiving and responding to a user's activity. We will consider a video interactive installation with hardware consisting of video input device (video camera), a processor (desktop or compact computing device) and a video output device (TV panel).

The key modules for interactive installation software are:

1. Video Analysis: interpretation of user's actions. It is possible to construct a robust computer vision module, working in uncontrolled light conditions and performing following tasks [2]:
   - detect objects appearing and disappearing;
   - detect objects' motion direction and speed;
   - recognize human face and its expression;
   - recognize several hand gestures.

   The output data of the module is enough for many camera-based interactive scenarios.
2. Behaviour & Dynamics: scene internal state representation and changing accordingly the user's actions.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 298–299, 2009.

3. Render: visual representation of the scene state. We propose to use image-based rendering [3] as the universal and simple way for utilizing the animation obtained with different methods: 2d, 3d animation and live video recordings. More specifically, it is proposed to represent the object's dynamics by a number of animation sequences. This sequences have common frames, in which it is possible to switch between the sequences while playing. Such sequences can be called "branching animation sequences" (Fig. 1).

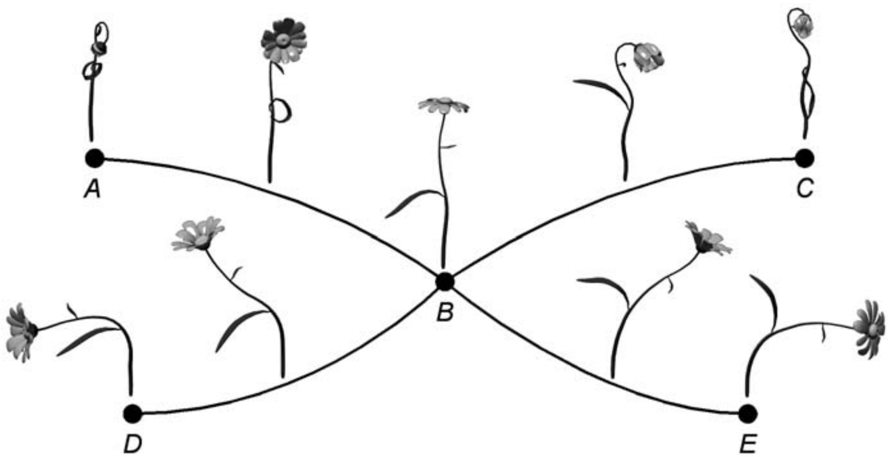


Fig. 1. Four Branching Animation Sequences

The experimental platform for testing the proposed technique was developed. A number of interactive installations were implemented. The experiments proved the simplicity of implementing the interactive scenes. It seems that the capabilities are quite enough for the mass production of the wide diversity of the simple interactive installations.

Part of the research was supported by the grant 09-01-00523 from the Russian Foundation for Basic Research and the Presidium of the Russian Academy of Sciences fundamental program 29, project P(29)7-2. The author would like to thank G. Malyshev and P. Zakrevskiy for 3d animation, S. Zamuraev for fruitful discussions, I. Ilyin for the help in programming, S. Perevalova for the help in demo, A. Poptsova for the text correction.

References

1. `http://vvvv.org`
2. Sebe, N., Lew, M.S.: Robust Computer Vision: Theory and Applications. Kluwer Academic Publishers, Dordrecht (2003)
3. Shum, H., Chan, S., Kang, S.B.: Image-Based Rendering. Springer, New York (2006)

Reinforcement Learning for Blackjack

Saqib A. Kakvi

Goldsmiths, University of London, SE14 6NW, London

Abstract. This paper explores the development of an Artificial Intelligence system for an already existing framework of card games, called SKCards, and the experimental results obtained from this. The current Artificial intelligence in the SKCards Blackjack is highly flawed. Reinforcement Learning was chosen as the method to be employed. Reinforcement Learning attempts to teach a computer certain actions, given certain states, based on past experience and numerical rewards gained. The agent either assigns values to states, or actions in states. This will initially be developed for Blackjack, with possible extensions to other games. Blackjack is one of the simpler games and the only current game in the SKCards package which needs an Artificial Intelligence agent. All the other games are single player. To test the performance of the Reinforcement Learning agent, several experiments were devised and run.

1 Background

1.1 Current System

The current system is a framework for all card games. There is a reresentation of Cards and gameplay is represented by the interdependant Interfaces Player and UI. The Blackjack Player has an attempted AI implementation, which is a number of nested if-else if-else clauses. This does not work.

1.2 Blackjack

[1] The object is to get upto 21 points in a maximum of 5 cards. This is achieved is by assigning values to each of the cards in the deck: 2-10=Face value; Courts=10; Ace=1/11. The player may then take another card(hit) or end thier turn(stand).

1.3 Reinforcement Learning

Reinforcement Learning learns from rewards for taking a sequence of actions in an evirnoment, based on its knowledge. This will lead to a change and eventually a reward. The agent will accordingly adjust its knowledge. Agents can act based upon values of states or actions in states. It learns based on past rewards for an action, A, and the reward received for action A. For simplicity, we take the average of the previous rewards and update it and also allows the agent to know how good a reward is.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 300–301, 2009.

The agent implemented is a softmax selection agent. It choses actions based on thier probability, from the equation below [2]:

$$\frac{e^{Q_t(a)\backslash\tau}}{\sum_{b=1}^{n} e^{Q_t(b)\backslash\tau}} \tag{1}$$

This strikes a balance between exploitation and exploration. The parameter $0 < \tau < 1$ is called the temperature and decides how much the agent explores.

The basic value method uses discounted. The agent passes back the reward multiplied by a constant $0 < \gamma < 1$, known as the discount rate. Multiplying the reward by the step-size parameter, $0 < \alpha < 1$, mitigates the effect of noise in the data. By calculating these values, the agent learns a policy, P, which is the action the agent will take in a given state. We aim to learn a policy $P \simeq P^*$.

2 Experiments and Results

4 basic tests were devised. Each test uses different values of rewards to achieve a different purpose. The first test uses the normal values: Win = +10; Push = +5; Lose = -10; Bust = -10. The subsequent tests involve doubling one reward value and keeping the rest at the norm. The dealer hits until a set value. The constants used were: $\tau = 0.5$; $\alpha = 0.6$; $\gamma = 0.75$; A sample of the results are given in Table 1.

Table 1. Final Results

DEALER POLICY	WIN REWARD	LOSE REWARD	BUST REWARD	FINAL POLICY	WIN (%)	LOSE (%)	BUST (%)	NET WINS(%)
11	10	-10	-10	17	44.62	17.6	30.66	-3.64
17	10	-10	-10	18	19.08	17.4	38.74	-12.86
11	20	-10	-10	16	45.98	23.1	22.02	0.86
12	20	-10	-10	17	42.76	19.54	29.82	-6.6
15	10	-10	-20	12	19.48	58.88	0	-26.26
16	10	-10	-20	12	18.3	58.26	0	-25.72
16	10	-20	-10	20	20.46	5.8	59.86	-31.12
17	10	-20	-10	20	16.14	6.04	59.02	-25.7

3 Conclusion

The results show that by altering the rewards, alters the policy of the agent. It is seen that altering negative rewards has a larger effect than altering positive rewards. From the net wins, we can see that policy alone cannot win.

References

1. Parlett, D.: Teach Yourself Card Games. Hodder Headline Plc., London (1994)
2. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge (1998)

"Plug: Secrets of the Museum": A Pervasive Game Taking Place in a Museum

Michel Simatic[1], Isabelle Astic[2], Coline Aunis[2], Annie Gentes[1], Aude Guyot-Mbodji[1], Camille Jutant[1], and Emmanuel Zaza[3]

[1] Institut TELECOM, 46 rue Barrault, 75634 Paris Cedex 13, France
gentes@telecom-paristech.fr, audeguyot@gmail.com, camillemontreal@gmail.com, michel.simatic@it-sudparis.eu

[2] Musée des arts et métiers, 292 rue Saint-Martin, 75141 Paris Cedex 03, France
{isabelle.astic,coline.aunis}@cnam.fr

[3] TETRAEDGE Games, 54 rue Marceau, 93100 Montreuil, France
ezaza@tetraedge.com

Abstract. "Plug: Secrets of the Museum" (PSM) is a game played with NFC-enabled mobile phones inside a museum containing dedicated passive RFID tags. During a PSM session, 8 teams exchange virtual cards representing objects located in the museum. These exchanges are done either with RFID tags or with other teams. PSM game design results in an educational and entertaining game which is much more attractive than the plain old treasure hunt proposed by several museums. Thus PSM is a good companion to discover and even take up a museum.

1 Introduction

Museums intend to arouse visitors' interest in their pieces. Computer games can contribute to this mission [1].

[2,3] present examples of such games. None of them meet simultaneously all of the following requirements: 1) the game is educational but also entertaining; 2) installation and operating costs are limited; 3) it can be played by anyone (whatever their age and their social category); 4) it is a family game; 5) the same player can play several times without getting bored.

This is why we have developed "Plug[1]: Secrets of the Museum" (PSM). Section 2 presents its game design and how it matches all of the requirements. Section 3 concludes.

2 Game Description

In PSM, 8 teams play with NFC-enabled handsets. They manipulate virtual playing cards, which represent specific objects of the museum. These cards are stored either on handsets, or on RFID tags located besides these objects.

[1] PLUG is a research project funded by *ANR* and labelled by *Cap Digital*. Its contributors are academics (*CNAM*, *Institut TELECOM*, *Université de La Rochelle*), companies (*Net Innovations*, *Orange*, *TETRAEDGE Games*), a museum (*Musée des arts et métiers*) and an association (*Dune–Aventure*).

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 302–303, 2009.

The goal is to get the best score throughout the game session. To gain points, the teams may prove: their collector's ability, by gathering on their handset 4 cards of the same family (through exchanges "card stored in their handset ⇔ card stored in a RFID tag" or "card stored in their handset ⇔ card stored in another teams handset"); their public-spiritedness, by storing a card to its reference RFID tag (thanks to an exchange "card stored in their handset ⇔ card stored in a RFID tag"); their generosity, by exchanging one of their cards with another team; their curiosity, by answering quiz related to the objects of the Museum.

PSM is educational as, in particular, players discover parts of the museum they would not have paid attention to, if they had not played PSM. Nevertheless players considered it as very entertaining.

Installation of PSM consists in the deployment of tags throughout the museum. Moreover no network is needed at game time. This results in limited installation and operating costs.

Because there are several ways of gaining points, PSM reaches a broad audience (*e.g.* young children like its collector aspect whereas seniors enjoy quiz).

PSM suits a family visit. Family members can choose to play either together (in the same team), or against each other (in different teams).

A player can play several sessions of PSM without getting bored: each time, they can decide to choose a different strategy (focus on generosity points, focus on curiosity points...). Moreover their competitors will always be different and this will give a different thrill to each session.

3 Conclusion

12 public game sessions (with 150 players) showed that PSM makes players take up the whole museum. PSM gives them the desire to stay after their session, in order to focus on some parts of the museum discovered during the game.

During *ICEC* 2009, we intend to demonstrate PSM by making participants play PSM sessions in the *Musée des arts et métiers* museum.

PSM key concept is to collect cards through exchanges with RFID tags or other teams. This concept can be generalized easily to any kind of museums, with limited installation and operating costs. Thanks to PSM's educational but also entertaining features, one of the main missions of museums (arouse visitors' interest in their contents) is fulfilled.

References

1. Gupta, A.: The End of the Game, a Mystery in Four Parts. Smithsonian.com (December 2008), http://www.smithsonianmag.com/arts-culture/The-End-Of-The-Game-A-Mystery-In-Four-Parts.html
2. Heumer, G., Gommlich, F., Jung, B., Müller, A.: Via Mineralia - a pervasive museum exploration game. In: Proc. of Pergames 2007, Salzburg, AT (2007)
3. Laurillau, Y., Paternó, F.: Supporting museum co-visits using mobile devices. In: Brewster, S., Dunlop, M.D. (eds.) Mobile HCI 2004. LNCS, vol. 3160, pp. 451–455. Springer, Heidelberg (2004)

In-Game Peer Performance Assessment Role That Fosters Metacognitive Agility and Reflection

Elaine M. Raybourn

* Sandia National Laboratories, P.O. Box 5800 MS 1188
Albuquerque, New Mexico USA, 87185
emraybo@sandia.gov

Abstract. In this paper we describe the development of a method and system for training metacognitive agility (self-awareness and self-regulated learning) in serious games applications. We introduce a unique design that features a novel role for real-time, in-game peer performance assessment and feedback to encourage user reflection and self-explanation. This approach has been implemented in two serious games currently in use today whose focus is intercultural competence and intercultural sensitivity education.

Keywords: serious game, metacognitive agility, reflection, in-game performance assessment, peer learning, intercultural competence, sensitivity.

1 Reflective Evaluator Role in Multiplayer Serious Games

A serious game is defined in this paper as the use of interactive digital technologies for training and education in private, public, government, and military sectors. In multiplayer serious games play is usually followed by a debriefing period in which key learning points and participant performance are discussed in a large group after an exciting game session. It is typical in these cases, therefore, for the reflection and peer learning to occur outside of the game experience, while the game experience focuses on individual task performance [1]. It is also typical to hear these users and instructors say that the real learning occurs outside of the game, not as part of the game. However, this trend need not persist. By applying principles from social learning theory and entertainment game design we can embed in-game opportunities for user real-time reflection and peer learning during game play.

Honing one's metacognitive agility and reflection is integral to competence education [1]. Metacognitive agility is the ability to actively control the learning process [2] such as possessing the ability to reflect and analyze the way oneself or others think, discern when different cognitive strategies are needed, and employ those strategies to enhance one's learning and performance [1]. Operating competently requires the

---

* Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 304–306, 2009.

ability to be aware of oneself and others, reflect on salient experiences, evaluate or assess situations, and act purposefully on those evaluations.

The reflective evaluator role is based on the Real-time In-Game Assessment, Evaluation and Feedback system [1, 3] which consists of a game interface and role for making in-game evaluations of player's actions, decisions, communications, etc. as they occur in real-time and as they correspond to competencies and learning objectives. In each multiplayer session, Trainees, instructors, evaluators, or subject matter experts play roles for characters or peer observer/evaluators.

Our design goal with the introduction of this new role for reflective learning (observation and evaluation) was to encourage growth toward intercultural sensitivity [4, 3] and cultural relativism [5], or the ability to reflect and evaluate phenomena relative to cultural context. Therefore our intent was to expand trainees' solution set for ill-defined problems through their game session and subsequent participation in a follow-up discussion. More detailed information on the role is available from [1, 3].

Peers in the reflective role help focus and teach others participating in the debriefing to connect theories and/or concepts to actions demonstrating stages of intercultural competence and sensitivity [3, 4]. Trainees in the reflective role are therefore held accountable for explaining criteria of intercultural competence to others after observing modeled behaviors. This often accelerates learning as they internalized concepts and new vocabulary more quickly than others.

By training in roles that allow users to act (character roles) and reflect (Reflective Evaluator role) the trainees perform different cognitive tasks. More concrete, active experimentation (e.g. negotiating from a different point of view) takes place with character role participation, while abstract conceptualization and reflection is fostered by the Reflective Evaluator role (e.g. pause, look at the problem in light of the cultural context, critically consider best practices, and communicate solutions to others). Thus intercultural competence education is grounded in experiential learning theory—namely concrete experience, active experimentation, reflective observation, and abstract conceptualization [6]. This approach is entertaining and educational. It is used with multicultural groups to engender out-of game discussions of cultural differences and diverse approaches to problem-solving in two serious games currently in use today. We are extending this work by conducting user studies that investigate how to further improve experiential learning and by adapting the game environment in real-time as a result of performance, reflection, and in-game peer evaluation.

References

1. Raybourn, E.M.: Training approaches for honing junior leader adaptive thinking, cultural awareness and metacognitive agility. In: Proc. I/ITSEC 2007, Interservice/ Industry Training, Simulation and Education Conference, Orlando, Florida, USA (2007)
2. Flavell, J.H.: Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist 34, 906–911 (1979)
3. Raybourn, E.M.: Applying simulation experience design methods to creating serious game-based adaptive training systems. Interacting with Computers 19, 207–214 (2007)

4. Bennett, M.J.: A developmental approach to training for intercultural sensitivity. International Journal of Intercultural Relations 10, 179–196 (1986)
5. Rogers, E.M., Steinfatt, T.M.: Intercultural Communication. Waveland Press, Inc., Prospect Heights (1999)
6. Kolb, D.A.: Experiential learning: Experience as the Source of Learning and Development. Prentice-Hall, NJ (1984)

Edutainment Games for Mobile Multimedia Museum Guidance Systems: A Classification Approach

Areti Damala

CEDRIC / CNAM,
282 rue Saint-Martin, Paris, France
areti.damala@gmail.com

Abstract. This paper proposes a first sketch on the state of the art regarding interactive edutainment games delivered through mobile multimedia museum guidance systems used in the context of a cultural visit. The goal is to identify current practices but also potential functional requirements, through the introduction of a first set of classification criteria emerging from a literature review of representative projects. The issue of related evaluation practices is also discussed.

Keywords: mobile learning, edutainment, mobile museum guides, games.

1 Introduction and Motivation for Research

About one third of museum and gallery audiences are made up by children accompanied by their families. As interactivity and action is vital for children, a perception that the museum visit will be fun is crucial. In this context, educational games for museum handheld devices may constitute an alternative to PDA-escorted guided visits, especially attractive for children and adults by triggering curiosity and promoting engagement physically, mentally and emotionally [1].

2 A First Set of Classification Criteria

Multimedia games for museum handheld devices can be solitary games (e.g. [2, 3]) or team games [4, 5, 6, 7]. Teams can be consisted of individuals [4, 5] or groups of two or more players [5, 6, 7]. When teams or team members are given the possibility to communicate between them, communication can be synchronous -as is more often the case- or asynchronous [2]. Another distinction can be made according to whether the game proposes the use of one or multiple delivery platforms [7] such as stationary workstations, video projectors or head mounted displays [4]. The environment in which the game takes place can also be used for classification purposes; three categories can be distinguished: outdoor games in cities, historical or archaeological sites [4], indoor games in museums and galleries [5, 6, 7] and games that combine tasks that should be accomplished both in an indoor and an outdoor environment [7]. Similarly, the games can also take place in a variety of computer environments: 2D environments, 3D environments, Augmented Reality environments or combinations of the above [7]. The examined games are often

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 307–308, 2009.

inspired by already well known educational museum games, like “treasure hunts” and observation games [4, 7] or mystery/detective games, in which the players have to solve a mystery case regarding one or several museum objects ([4, 5]). According to the nature of the proposed activities, edutainment applications for mobile multimedia museum guidance systems can be also distinguished in observation games, reflection games or video/arcade games [2]. Finally it should be noted that additional activities are sometimes also proposed for the pre- and post-visit phases [3].

3 Perspectives and Directions for Future Work

Games can provide an important ally in mobile museum guides’ and edutainment projects. However, the number of evaluation studies carried out regarding edutainment interactive applications on museum handheld devices is significantly inferior to this concerning the effectiveness of use of mobile multimedia museum guides. In addition, scarce are the studies regarding the effectiveness of PDA delivered educational games as compared with more “traditional”, educational, paper and pencil quests. Hence, the urge to further invest in the conception, assessment and evaluation of mobile museum guides’ edutainment applications is crucial.

References

1. Damala, A.: Design Principles for Mobile Museum Guides Using Visitor Studies and Museum Learning Theories. In: Proceedings of the 2007 IADIS M-Learn Conference (Mobile Learning), Lisbon, Portugal, pp. 277–281 (2007)
2. Belotti, F., Berta, R., De Gloria, A., Margarone, M.: User Testing a hypermedia tour guide. Pervasive Computing Journal 1(2), 33–41 (2004)
3. Sauer, S., Goebel, S.: Dinohunter: Game based learn experience in Museums. In: ICHIM 2003, non-paginated document. Archives and Museum Informatics Europe, Paris (2003)
4. Hall, T., Ciolfi, L., Bannon, L., Fraser, M., Benford, S., Bowers, J., Greenhalgh, C., Hellström, S., Izadi, S., Schnädelbach, H., Flintham, M.: The Visitor as Virtual Archaeologist: Explorations in Mixed Reality Technology to Enhance Educational and Social Interaction in the Museum. In: Virtual Reality, Archeology, and Cultural Heritage 2001, pp. 91–96. ACM Press, New York (2001)
5. Broadbent, J., Marti, P.: Location Aware Mobile Interactive Guides: Usability Issues. In: International Cultural Heritage Informatics Meeting ICHIM 1997, pp. 15–30. Archives and Museum Informatics Europe, Milano (1997)
6. Thom-Santelli, J., Boehner, K., Gay, G., Hembrooke, H.: Beyond just the facts: transforming the museum learning experience. In: CHI 2006, Human Factors in Computing Systems, pp. 3214–3222. ACM Press, New York (2006)
7. Laurillau, Y.,釣Paternó, F.: Supporting Museum Co-visits Using Mobile Devices. In: Brewster, S., Dunlop, M.D. (eds.) Mobile HCI 2004. LNCS, vol. 3160, pp. 451–455. Springer, Heidelberg (2004)

Orpheus: Automatic Composition System Considering Prosody of Japanese Lyrics

Satoru Fukayama[1], Kei Nakatsuma[1], Shinji Sako[2], Yuichiro Yonebayashi[1], Tae Hun Kim[1], Si Wei Qin[1], Takuho Nakano[1], Takuya Nishimoto[1], and Shigeki Sagayama[1]

[1] The University of Tokyo
{fukayama,k-nakatsuma,yonebayashi,kim,qin,t-nakano, nishi,sagayama}@hil.t.u-tokyo.ac.jp
[2] Nagoya Institute of Technology,
sako@mmsp.nitech.ac.jp

1 Song Composition Algorithm Considering Prosody

We present an algorithm for song composition using prosody of Japanese lyrics. Since Japanese is a "pitch accent" language, listener's apprehension is strongly affected by the pitch motions of the speaker. For example, the meaning of Japanese word "ha-shi" changes with the pitch. It means "bridge" with an upward pitch motion, and "chopsticks" with the motion inversed. A melody attached to the lyrics cause an effect similar to the pitch accent. Therefore we can assume that pitches of Japanese lyrics give constraints on pitch motions of the melody. Furthermore, chord progression, rhythm and accompaniment give constraints on the transitions and occurrences of the melody notes. If a certain melody for the lyrics were obtained, the melody would satisfy these constraints. Conversely, we can compose a song by finding the melody which optimally meets the condition.

2 Implementation and Experimental Results

Orpheus is an automatic composition system that we implemented using melody composition algorithm based on prosody. This system computes melody from the lyrics input with choices of chord progressions, rhythm patterns, and accompaniment instruments. We used Galatea-Talk[4] text-to-speech engine to analyze the prosody of Japanese lyrics, and HMM singing voice synthesizer[5] to generate the vocal part. We also implemented the system as a web-based application[1].

We did two experiments to evaluate the system. Firstly, we asked a classical music composer to evaluate 59 generated songs in five-grade evaluation. Secondly, we uploaded our system to get comments from a large number of users on the internet. During a year of operation, about 56,000 songs were generated by the users and 1378 people answered the questions about *Orpheus* and the generated songs. The results are shown in Fig. 1 and Fig. 2. Judging from the results, about 70.8% commented that the generated songs are attractive, and 84.9% of the users had fun trying this system.

[1] http://orpheus.hil.t.u-tokyo.ac.jp/automatic-composition/index.en.cgi

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 309–310, 2009.

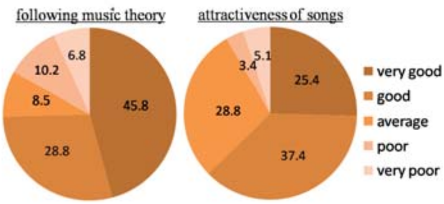


Fig. 1. Evaluation results on 59 songs by a classical music composer [%]

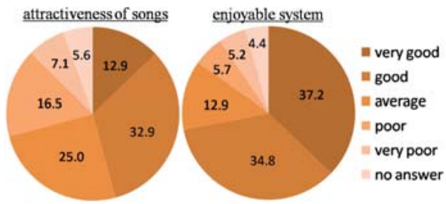


Fig. 2. Evaluation results on generated songs and the *Orpheus* by 1378 users [%]

3 Discussions and Conclusion

This research attempted to design an algorithm to compose a song automatically from the lyrics using prosody information, which enables users to make their original songs easily. The results indicate that our method and implemented system *Orpheus* is an enjoyable solution for amateur musicians.

However, it should be noted that our algorithm can be applied to lyrics written in "pitch accent" languages only. As a next step, we plan to extend the composition algorithm to handle "stress accent" languages, such as English, by putting constraints on metric structure of the melody.

Acknowledgments. We would like to thank Hitomi Kaneko for the evaluation of generated songs and useful discussions from the composer's point of view. This research was funded by the Crest Muse Project[6].

References

1. Hiller, L., Isaacson, L.: Experimental Music. McGraw-Hill, New York (1959)
2. Xenakis, I.: Formalized Music, Revised edn. Pendragon Press, New York (1992)
3. Cope, D.: Computers and Musical Style. A-R Editions. Madison, Wisconsin (1991)
4. Galatea Project, `http://hil.t.u-tokyo.ac.jp/~galatea/`
5. Sako, S., Miyajima, C., Tokuda, K., Kitamura, T.: A Singing Voice Synthesis System Based on Hidden Markov Model. Transactions of IPSJ, 719–727 (2004) (in Japanese)
6. CrestMuse Project, `http://www.crestmuse.jp/index-e.html`

A Handy Laser Show System for Open Space Entertainment

Toru Takahashi[1], Miki Namatame[2], Fusako Kusunoki[3], Isao Ono[1], and Takao Terano[1]

[1] Tokyo Institute of Technology, Nagatuda, Midori-ku, Yokohama, Japan
{toru@trn,isao@,terano@}.dis.titech.ac.jp
[2] Tsukuba Institute of Technology, Tsukuba, Ibaragi, Japan
miki@a.tsukuba-tech.ac.jp
[3] Tama Art Univeristy, Hachioji, Tokyo, Japan
kusunoki@tamabi.ac.jp

Abstract. Big Fat Wand (BFW) is a handy laser show system, which includes a portable laser show device newly developed and a laptop PC with easy-to-use authoring toolkits. This paper describes basic principles and architecture of BFW, then demonstrates how BFW is used in an open space environment.

Keywords: Laser Show Device, Entertainment in an Open Space, Interactive Sessions.

To enhance human-computer interaction activities, we are developing a new intelligent pointing device: Big Fat Wand (BFW). Compared with a conventional laser show device [1], the size is as small enough as we use it in hand. Furthermore, compared with a conventional laser pointer, it is connected with a laptop PC and programmable to allow the user to specify the pattern displayed via any characters and/or symbolic patters on the targeted object. BFW stands for the very big magic wand. The device displays various information to any kinds of target objects, even if they are in a bright place.

BFW originally aims at edutaining hearing-impaired students for out-of-classroom lectures. So far, intensive experiments on art sculpture lectures at a large room have uncovered that BFW is a cutting-edge system for the purpose [2]. However, BFW can be used as a general entertainment tool.

Big Fat Wand system has the following components: i) A laptop PC with line drawing image generation, image display, and editing software, ii) A one-board micro-computer to convert the digital information of the drawings to the analog ones to control the device, and iii) Laser show device with laser light generator, small dynamic mirror devices to control displays of the drawings, and power supplies. The system is also equipped with special purpose authoring tools for naïve users to prepare the explanation materials.

Very unique points of BFW are summarized as follows: i) A one-board 16 bit micro-computer manages DA conversion of explanation and controlling the images, ii) The portable cylinder part is carefully designed to avoid heat damages of the laser devices, and iii) the components of the devices are packaged in separated two parts to easily use the system.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 311–312, 2009.

We show how BFW is used for entertainment purposes. The first example is to display any images into target objects (Figure 1). The second example is about an on-going project *Skelton Dances*, whose objective is to develop a new interactive entertainment system with both displayed animated characters and human actors (Figure 2).

Our future work includes 1) to prepare interesting interactive scenarios for the display images and human actors and 2) to improve BFW to Small Smart Wand.

Fig. 1. Demonstration of BFW for Daily Explanation and Amusement

Fig. 2. Interactive Session with a Display Image and a Human Actor

References

1. Graffiti Research Lab (2009), `http://graffitiresearchlab.com/`
2. Takahashi, T., Namatame, M., Kusunoki, F., Terano, T.: Big Fat Wand: A Pointing Device for Open Space Edutainment. In: Proc. INTETAIN 2009 (to appear, 2009)

Sketch-It-Up! Demo

Bulut Karakaya, Camilo Garcia, Daniel Rodriguez, Manoj Nityanandam,
Nadia Labeikovsky, and Theyab Al Tamimi

700 Technology Drive, Pittsburgh, PA – 15219, United States of America
contact@sketch-it-up.com

Keywords: Ideation, Risk free innovation.

1 Description

Every creative project needs to have an ideation process. A good ideation process relies on a simple yet effective way of putting ideas on the table and sorting through them, also discarding them easily if necessary. However, for this ideation to be successful it has to be a process in which cheap and simple ways of exploring ideas are used along with tools that are readily accessible to everyone in the field to use.

It is known that the video game industry is mostly a production-oriented industry. This emphasis on production has neglected research into the ideation process necessary in a field of this nature. Very little has been taken into account regarding new ideas of gameplay, rather the focus is more on the aesthetics of the game. This means that the gameplay of many games of the same genre is identical, except for the background story and characters. This phenomenon is linked to the rapid growth of the industry and the vast sums of money that it is needed in order to produce a modern game.

We feel that this is a problem that should be addressed and solved, and for this we have suggested a tool in which ideas can be "sketched" in minutes with a whole team of designers and developers.

This tool is called Sketch-it-up!, It is an extension of the GameSketching system conceived by Dr. John Buchanan and his team in 2007. It is not just a software tool or a solution to a particular problem, but rather a set of processes and technologies that allow game development teams to explore and communicate ideas. Sketch-it-up! builds on the previous GameSketching system making it more accessible, whether the users are gamers, programmers, designers or artists.

Sketch-it-up! is a process that should happen in parallel to ideation, brainstorming and tools creation in games. Not to be confused with *prototyping*, which would require having game ready assets, the concept of *sketching* relates to hasty executions of ideas without using game ready assets or code. This would enable focused conversation between members of a team about the bare bones of a game, the essentials, the pros and cons of the idea itself.

Up to this day this ideation process was done through paper sketching and similar contraptions. Our tool is different because it is interactive in such a way that every member of a team can control different aspects of the sketch. It also reduces the confusion on writing and erasing things in the same paper at once. Different views are

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 313–314, 2009.

also provided so that the idea can be expressed from different angles and different game genres can be sketched. It can be shared and worked upon in real time on any network like ad-hoc, LAN or internet which is a huge advantage over paper sketching.

Sketch-it-up! is built on a Client/Server architecture. We refer to the Clients as *puppeteers,* and these puppeteers are connected to a main server. The puppeteer and server then load up the same world and each time an object is added by a puppeteer, it appears throughout the world. To avoid confusion, all requests to add/modify/move characters go through the controller and are granted only if those objects belong to that particular puppeteer. Each of the puppeteers has his own camera view to facilitate better maneuvering of his puppets.

Sketch-It-Up! is built on Disney's Panda3D engine because of its simple licensing policy and rapid development cycle due to the use of Python. Sketch-It-Up! is preloaded with a number of assets that can be used to sketch many types of games. But if needed, assets can be created in Maya, 3D Studio Max or Blender and imported into the tool.

To make a tool that the industry really needs, we presented it to game companies and got their feedback. We also constantly tested the tool ourselves to fix bugs and add features. This led us to make the current, robust tool that we have now.

However, in making this tool, we found that it has its limitations, as not every question or every genre can be sketched. But we see that features can be born during a sketch thanks to the spontaneity and improvisation of the puppeteers doing the sketch. Situations can be created in order to reinforce a game idea already implemented making that game unique even though its genre relies on a determined mechanic, such as fighting games like Street Fighter where we could sketch the layout of the fight, but not the fight itself.

We have found that sketching also improves the quality of ideas while saving time and money while enabling the team to have fun in the process.

References

1. Sketch-It-Up!, `http://www.sketch-it-up.com`
2. Panda3D, `http://www.panda3d.org`

Automatic Chat Generation of Emotional Entertainment Characters Using News Information

Jun'ichi Hoshino, Tetsuya Saito, and Kenichi Hirota

University of Tsukuba, Graduate school of Systems and Information Engineering,
1-1-1, Tennodai, Tsukuba-shi, Ibaraki, Japan
{jhoshino,stetsuya,hirota}@entcomp.esys.tsukuba.ac.jp

Abstract. Currently interactive entertainment characters have their conversation topics prepared beforehand by the creator. In this paper, we propose the automatic chat generation engine for emotional entertainment characters using real-time news information. The character emotionally reacts to the news contents based on their interest and feelings so that the users can have intimate feeling. The character also incrementally learns user's interest from their response.

Keywords: chat generation, RSS, emotional character, user learning.

1 Introduction

Currently, various games and movies offer the means to conduct conversations with characters. In role-play games, there are often scenes in which a player talks about the progression of the game with non-player characters, hears two characters talking, or joins in the small talk. In this paper, we propose an entertainment character that generates various utterance topics using dynamic sources of news, weather reports, etc., which exist on the Web as sources of real-world information. Since the Web information used is up-to-date, the utterance content of the character can be updated at any time without the need for the creator's intervention. Moreover, an expression of feeling for the topic is a critical factor in improving the expressivity of the utterance, and in this paper, we propose a technique for simply generating the expression of the character's feeling from the directionality of the chat content. Using this technique, the character can express surface feelings appropriate to the topic. The latest news, weather forecasts, etc., are used as chat topics so that generating an appropriate conversation becomes possible using information acquired from the Web.

2 Related Works

Studies dealing with a dynamic source involve a sport system based on ball movement, game data, and the team impression in soccer simulation [1]. However, the focus is on an ongoing soccer game, and therefore, not suitable for utterances in chat communications. In addition, studies involving characters that read Web content, convert it into broadcasts and talk-show type content using character animation

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 315–318, 2009.

[2][3][4]. Interfaces that interactively retrieve a formula from the user and some characters have also been proposed [5][6].

3 System Design

Fig. 1 shows the outline chart of the technique. This system talks about the news and weather as real-world information. The topic is acquired and accumulated from XML and RDF files describing the news, weather forecasts, etc., on the Web, which are delivered as RSS feeds. Hereafter, these data files have been designated Web information files.

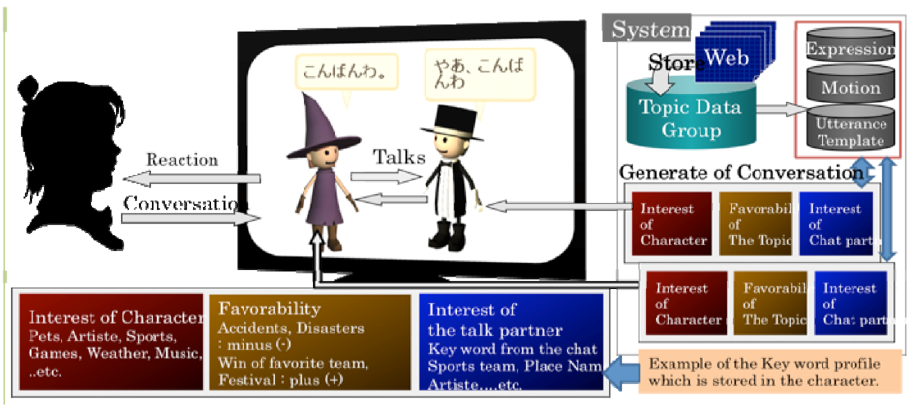


Fig. 1. Outline of the chat generation system

Conversation between the user and the character starts with text that the user inputs or with the character's voluntary utterances to the user. The topic is decided using the user's and character's level of interest to generate a significant utterance for the user, and not just make random conversation. If a topic garners a high level of interest, it can easily be selected as the conversation content. Moreover, the level of interest in the topic changes tending to favor the user and the character through small talk.

This system uses parameters of favorability that judge the topics as pleasant or unpleasant to improve the character's utterances. When the character utters, it automatically calculates the favorability of the topic and selects an appropriate expression, behavior, and the utterance template.

The conversation flow is coordinated according to several patterns and used to enable continuous utterances and an easy conversation between the user and the character. Moreover, the flow of an appropriate conversation between characters is achieved by equipping the character with a conversation pattern that has a level of interest similar to that of the user.

When the conversation begins from the user's text input, the character selects a topic suitable for the user's demand. The input content is accumulated as an information genre in line with the user's interest. When the conversation begins from voluntary utterances made by the character, the topic is selected from the genre in which the character has an interest or that in which the user has an interest. Next, expressions, gestures, and the utterance template of the character that match the favorability are selected by referring to the calculated favorability of each topic resulting in an animated utterance.

Fig. 2 shows an example of an expression and operating the character according to favorability. Since these operations consider only the directionality of the content in the chatting topic, the character does not develop feelings that the person actually has beforehand. However, we consider that it is superficially possible to react appropriately for the topic.

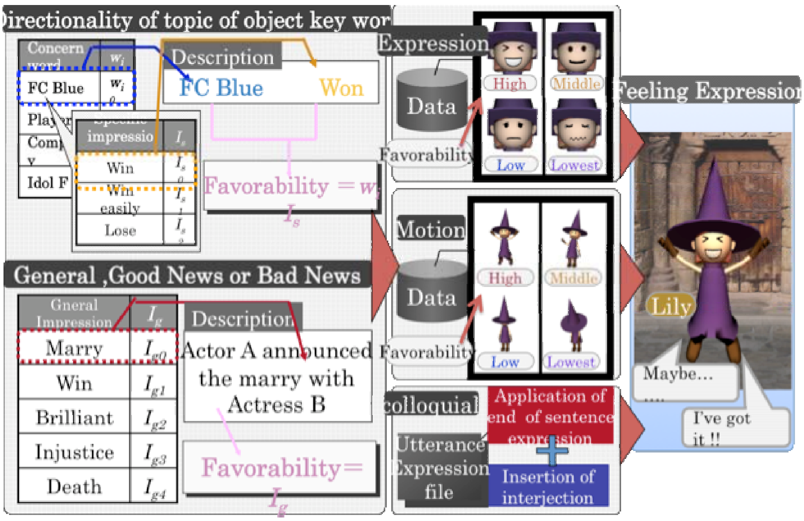


Fig. 2. Example of Feelings and Gestures for the Topic

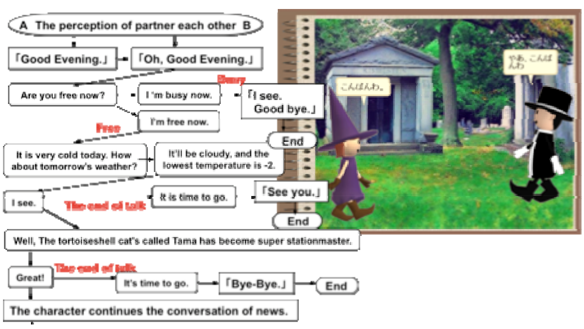


Fig. 3. Example of Conversation between Characters

4 Conclusion

In this paper, we propose the concept of emotional chat for characters in games through the use of Web information to generate utterances. In the proposed system, the conversation character can behave more naturally toward the user and talk flexibly. In the future, we intend to develop this system and obtain natural-speaking characters that can engage in interesting conversations.

References

1. Andre, E., Binsted, K., Tanaka-Ishii, K., Luke, S., Herzog, G., Rist, T.: ”Three RoboCup simulation league commentator systems. Artificial Intelligence Magazine (21), 73–85 (Spring 2000)
2. Akiyo, N., Taeko, H., Hiroyuki, K., Ikuo, S., Katsumi, T.: Automatic Conversion and a Scripting Markup Language for Passive Watching and Listening of the Web Contents. Transactions of Information Processing Society of Japan 42(SIG1) (TOD8), 103–116 (2001)
3. Akiyo, N., Katsumi, T.: Passive Viewing of Web Page based on Automatic Daialog Generation. IPSJ SIG Notes 2004(72) 2004-DBS -134(I), 183–190 (2004); Cavazza, M., Charles, F., Mead, S.J.: AI-based Animation for Interactive Storytelling. In: Proceedings of Computer Animation, pp. 113–120. IEEE Computer Society Press, Seoul (2001)
4. Hourai, H., Akiyo, N., Katsumi, T.: Transformation from Web contents into humorous user-friendly talk show type contents. DBSJ Letters 2(2), 29–32 (2003)
5. Kitamura, Y., Yamada, T., Kokubo, T., Mawarimichi, Y., Yamamoto, T., Ishida, T.: Interactive integration of information agents on the web. In: Klusch, M., Zambonelli, F. (eds.) CIA 2001. LNCS, vol. 2182, pp. 1–13. Springer, Heidelberg (2001)
6. Kitamura, Y., Sakamoto, T., Tatsumi, S.: A competitive information recommendation system and its behavior. In: Klusch, M., Ossowski, S., Shehory, O. (eds.) CIA 2002. LNCS, vol. 2446, pp. 138–151. Springer, Heidelberg (2002)
7. Mehrabian, A.: Communication without words. Psychological Today 2, 53–55 (1968)

Incremental Learning Algorithm for Online Action Game System

Junichi Hoshino and Hiroshi Mori

University of Tsukuba, Graduate School of Systems and Information Engineering,
1-1-1, Tennodai, Tsukuba-shi, Ibaraki, Japan
{jhoshino,hmori}@esys.tsukuba.ac.jp

Abstract. One of the limitations of computer opponents in action games is that the character AI is constructed in advance, and players may become bored quickly. We have built an online action game system in which a non-player character (NPC) can incrementally learn sequences of action and combinations. NPCs can adopt different fighting strategies for fighting with different players.

Keywords: action game, imitation learning, non-player character.

1 Introduction

One of the limitations of computer opponents in action games is that the character AI is constructed in advance, and players may become bored quickly. A human player can play the same game repeatedly, learn the behavior of the computer-controlled game character, and win easily. This is one major reason why human players soon grow tired of "fighting with a computer." Thus, many players prefer playing with another human player to playing with a computer.

In this paper, we propose an online action game system in which a non-player character (NPC) can incrementally learn sequences of action and combinations in real time. NPCs can adopt different fighting strategies for fighting with different players. Individual fighting styles can be generated from the unique fighting history. We have developed a new engine of action learning that analyzes a human player's action pattern automatically and extracts the effective fighting sequences. Action control trees are generated automatically and incrementally added to a player's action profile.

2 Learning Game Character

Fig. 1 shows the concept of an online colosseum game system in which non-player characters (NPCs) can incrementally learn sequences of action and combinations by fighting with human players. NPCs can adopt different fighting strategies for fighting with different players. Fig. 2 shows the growing process of NPCs. Individual fighting styles can be generated from the unique fighting history.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 319–322, 2009.

A player profile is created from the play data of the imitated player. The information recorded in a player profile includes the player's tactics, which we discussed in the preceding sections, the tactic sequences that represent the player's strategies, and the frequency with which each of these sequences appears. When the system applies a profile to the actual playing of the game, it reads the tactics and strategies and creates a chart of the tactic graph based on the tactic sequences (strategies) and stores it in the pool of graphs. While playing a game, the computer chooses a tactic graph that matches the selection criteria closest to the situation at a specific time, chooses the tactics in the tactic graph that corresponds to the situation that such a tactic graph refers to, and applies the actions to the game character by referring to these tactics. In the following sections we will discuss how the tactics and strategies are chosen.

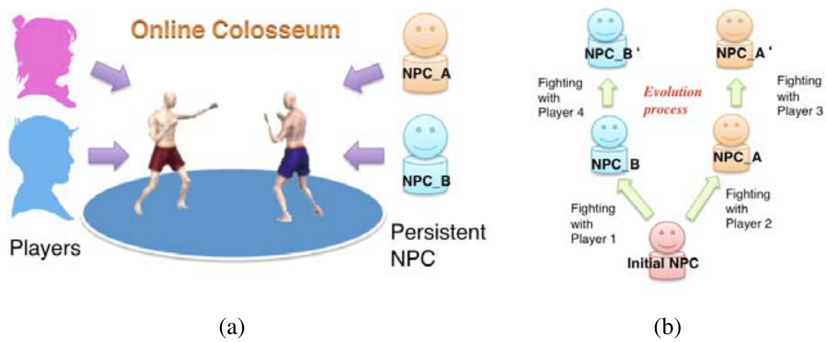


Fig. 1. Concept of an online action game system. (a)Many players and persistent NPCs can fight each other on the virtual colosseum. (b)NPCs can obtain different skills by fighting with different players.

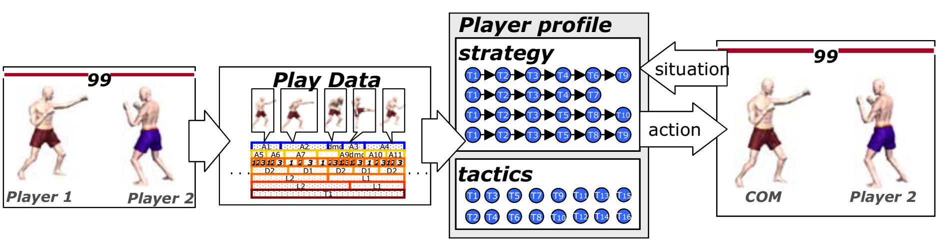


Fig. 2. Overview of the player profile sytem

3 Learning Strategy from a Human Player

1) Common Sequences of Tactics

Searching for a usable action pattern from accumulated action patterns is a technique often used for creating fighting strategies. In this paper, this is called method 1. The sequences of tactics in the player profile, created from the game data, are evaluated, and patterns that the opponent used repeatedly are given a high value. However, if a tactic is used too much, it might not be imitated. A certain value is chosen that represents the ratio of that action to the total length of the sequence of tactics. The standard for what sequence of tactics is suitable to imitate changes depending on this value. Therefore, setting of this at a different in various characters affects how the character evolves.

2) The Sequence of Tactics Contains Actions That the NPC Executes

This method of imitating adversary tactics includes using actions that we execute effectively. In this paper, this is called estimation method 2. In this method, new actions that the NPC does not execute are not added, because only actions that the NPC has already used are evaluated. This evaluation adds new conditions for executing actions that have already been used, and new sequences of tactics composed of existing actions.

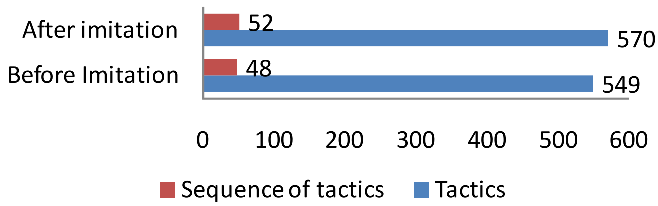
3) Sequences Containing Actions that the NPC Does Not Execute

Imitating the tactics of an adversary that do not include actions that we execute effectively is a method used when a person imitates how to fight others. In this paper, we call this estimation method 3. In this method, the new actions—ones that the NPC does not execute, are added by evaluating sequences consisting only of these actions. Therefore, adding these sequences enables the NPC to execute new actions. The evaluation value is determined from the success rate, the occurrence count, and the sequence length, as in the previous evaluation method.

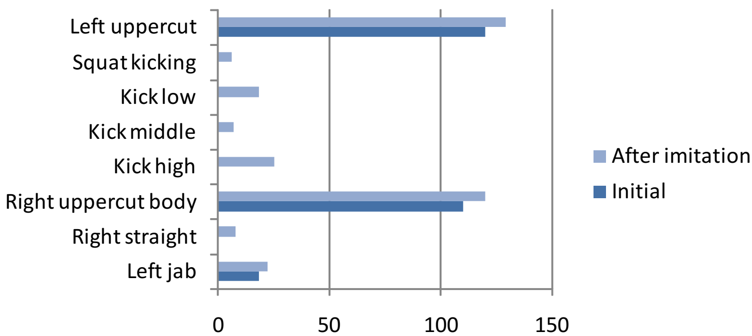
4 Experiments

Fig. 3 shows a screen shot of our fighting game. The initial player profile was made from the play data of 10 games. This initial player profile contains 549 tactics and 48 sequences of tactics. A player played ten games against the NPC based on this player profile, and we checked the added tactics and sequences of tactics as a result of the ten games. The probability of imitating the adversary's sequence was set high because it was an evaluation for only 10 games. As shown in Fig 4(a), four sequences and 21 tactics were added after 10 games. Fig 7(b) shows the number of the major actions.

Fig. 3. Screen shots of the fighting game



(a)The number of tactics and sequence of tactics



(b) The number of the major actions.

Fig. 4. Result of imitation learning

5 Conclusion

In this paper, we proposed a system that enables a computer character to imitate a human player. To do so, the system first acquires tactics and tactic sequences from play data of a player. Then, from the tactic sequences collected it creates tactic graphs that represent the strategic actions of the player. From these graphs, the system selects tactics that suit different situations. We also demonstrated the effectiveness of the system in an evaluation experiment. Furthermore, we created many different behavioral patterns for the computer by changing player profiles, which are the collections of tactic sequences and tactic graphs of the particular players.

Task-Based Second Language Learning Game System

Jun'ichi Hoshino, Tetsuya Saito, and Shiratori Kazuto

University of Tsukuba, Graduate school of Systems and Information Engineering,
1-1-1, Tennodai, Tsukuba-shi, Ibaraki, Japan
stetsuya@entcomp.esys.tsukuba.ac.jp, jhoshino@esys.tsukuba.ac.jp

Abstract. In traditional English learning as a second language, learners rarely have the opportunity to practice oral communication, so the acquisition of oral proficiency is a slow process. In this paper, we propose a task-based second language learning game system. The task-based learning method enables learners to obtain communicative skills through the practice of particular "missions" using voice and gesture communications with life-size 3D game character.

Keywords: Task-based language training, Conversational game character, Locomotion control, Negotiation of meaning.

1 Introduction

In traditional English learning in Japan, learners rarely have the opportunity to practice oral communication, so the acquisition of oral proficiency is a slow process. On the other hand, the task-based learning method enables learners to obtain communicative skills through the practice of particular "missions". As a result, the efficient improvement of communication skill becomes achievable. In this paper, we propose a task-based second language learning game system. The task-based learning method enables learners to obtain communicative skills through the practice of particular missions using voice and gesture communications with life-size 3D game character.

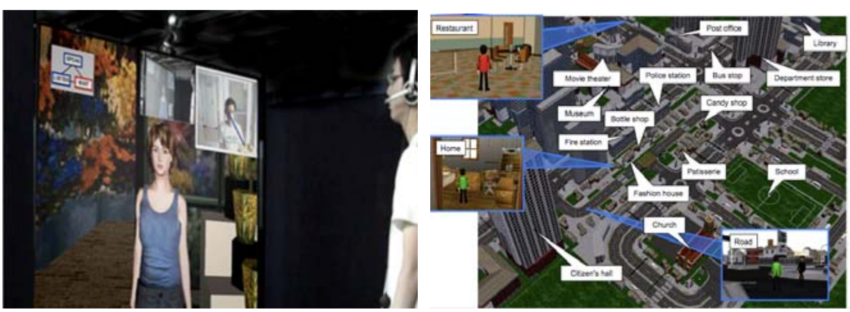


Fig. 1. A snapshot of using second language learning game system and the map of the town. Many daily conversational tasks can be experiences.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 323–324, 2009.

2 Game System Design

Fig. 1 shows a snapshot of using second language learning game system. The use environment of the system is a public classroom or a personal study room. The learners are assumed to be people who studied English in junior high school and high school. As a result, further study should build upon the skills one has learned, instead of wasting the knowledge acquired in school. The interface of the system consists of a large screen, a camera for the user behavior analysis, and a voice recognition system. The game character recognize user's voice and return conversation sentences based on the negotiation of meaning model.

The learner experiences task-based language training in various daily situations such as shopping, cooking, and job. The task-based learning method enables learners to obtain communicative skills through the practice of particular missions using voice and gesture communications with life-size 3D game character.

This system, using TBLT methodology, provides the negotiation of meaning function, and through its use, the imperfect understanding between the learner and the game character can be resolved. The learner advances through tasks by 'talking' with English-speaking characters in a virtual city environment. The content of the conversation is displayed in a balloon above the heads of the characters, and the learner types their response.

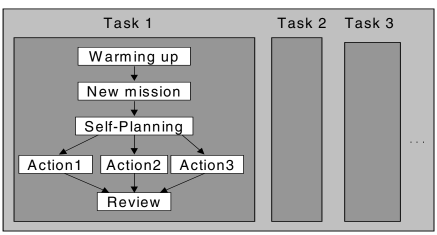


Fig. 2. Task model for TBLT

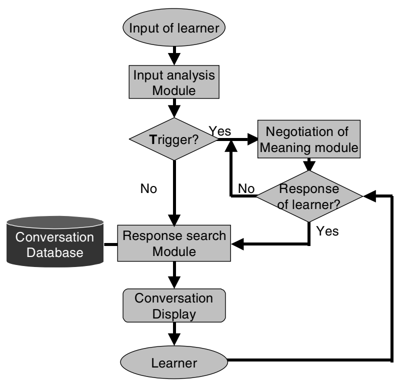


Fig. 3. Flow chart of conversation

3 Conclusions

We have presented an approach to promoting communication and achieved learner's greediness for learning through the negotiation of meaning model in the task-based learning system. Moreover, the effectiveness of the game character's was acknowledged in the learner's evaluations of the experiment. In future, it will be necessary to evaluate a series of tasks in order to improve the conversation function of game character's.

Designing a Game Controller for Novice HALO3 Players

Matthijs Kwak and Ben Salem

Department of Industrial Design, Eindhoven University of Technology, Den Dolech 2, 5600MB, Eindhoven, The Netherlands
matthijs.kwak@gmail.com, mail@bsalem.info

Abstract. This paper describes the process of designing and developing an intuitive controller that helps lower the threshold for novice gamers to play Halo 3. To help novices to master the controller, most controls have been replaced with rich and meaningful interaction. To help novices understand the game, extra feedback channels have been added to reinforce information given onscreen so critical information is not missed.

Keywords: Game Controller, HALO3, XBOX 360, and Product Design.

1 Introduction

The most important objective of this project is to lower the threshold for novices to play games. While the design of the current XBOX 360 controller is not difficult for novices to hold, the controls (buttons, joysticks) require a lot of training to master them, which is often good enough reason for novices to stay away from the XBOX 360. We are also concerned with providing some guidance for novice players. The result of all this is consisting of two controllers, one for each hand, providing the user with more freedom. These controllers are equal in level of importance, sharing the same design and dividing the controls as evenly as possible. In the redesign of the controller, we are focusing on some key feedback given to players (see the work of Paras (2006) in this area).

2 Aiming

Aiming has specifically received a lot of attention. It is done by changing the angle of your head and turning around your axis. We have implemented a control that combines the strengths of the joystick (relative control) and the mouse (absolute control).

To give a more natural link between the control and the activity (aiming), a slider, equipped with force sensing resistors (FSRs) is used. By sliding the slider with the thumb the angle of the characters head is determined in an absolute way. By applying force with that same thumb, either to the right or the left of the middle, the acceleration by which the character turns around his axis is determined in a relative way. We have evaluated our design with user tests, focusing on the novelty of our design and how easy can one learn to use the controller. We asked 18 students at Industrial Design to use three devices in random order to perform a targeting task on a computer display.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 325–326, 2009.

Figure 1 shows the improvement of the scores (means) for the left and right hands for the three devices: Joystick (1), Mouse (2), Combined Controller (3) after 5 repeated sessions with each device. While for the left hand, the Mouse is the best improver, for the right hand it is our combined controller that has the best improvement.

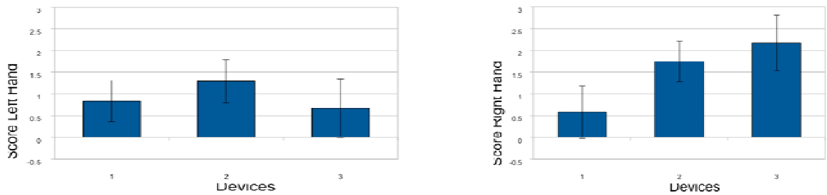


Fig. 1. Score improvement (means) for Joystick (1), Mouse (2) and combined controller (3)

4 Conclusion

We have designed, developed and integrated several original controls and coupled interactions within our controller. Novices should have less difficulty in understanding the game mechanics and more fun playing Halo 3. Our design lets gamers control the Halo 3 character without having to displace his/her fingers and thumbs. The device has controls and interactions that are linked more naturally to the outcomes in game and placed more logically. Furthermore, the guidance in the controller enforces the information given on screen. The reloading prompt, enacted by the dropping of part of the controller, is very strong. It is also a combination with the reload control.

Reference

Paras, S.P.: Learning to Play: The Design of In-Game Training to Enhance Videogame Experience. PhD thesis. Simon Fraser University Surrey (2006)

AZ66: How Can We Play with Emotions?

Stéphan Froment[1], Mélanie Ginibre[1], Stéphanie Mader[1], Antoine Sarafian[1], Aymeric Schwartz[1], Delphine Soriano[1], Alexandre Topol[2], and Jérôme Dupire[2]

[1] ENJMIN – 121, rue de Bordeaux
16000 Angoulème, France
[2] CNAM - CEDRIC – 292, rue St Martin
75003 Paris, France
{topol,dupire}@cnam.fr

Abstract. Since the early 90's, affective computing researches have mainly been focused on emotions recognition and less on original entertainment contents that could be proposed. We present a prototype resulting from the association of a physiological sensing device with an original game.

Keywords: video game, affective interaction, physiological sensors.

1 Related Works

Since the precursor work of Picard in the 90's [1] many researches were conducted in the affective computing field. First of all, physiological aspects were studied to determine which variables are interesting to be monitored. Then, patterns for recognizing specific emotions were revealed by looking to variables' variations in particular situations. Since then, the hardware needed to monitor humans became smaller, more accurate and even wireless [2]. The researches led to numerous fields of application, from health care [5] to entertainment computing [6, 7].

Using the physiological information to disturb the game, adapt its contents and even manipulate the player is quite interesting, in a game validation process. It would be very useful for a game development team to be able to test some particular aspects of the gameplay, from the early development stages. We present in this paper such a game (prototype) born from the association of a physiological sensing device with an original virtual reality game.

2 AZ66

The physiological monitoring device was built by us [8]. We designed a hardware which is able to monitor the heart signal (EKG), the galvanic skin response (GSR) and the temperature (TEMP). Thanks to a socket mechanism (Fig.1), all the data is transmitted wirelessly and integrated into the Unreal Engine 3. The interactions remain typical (through the keyboard and the mouse) and only the monitoring hardware differs from a usual playing environment.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 327–328, 2009.

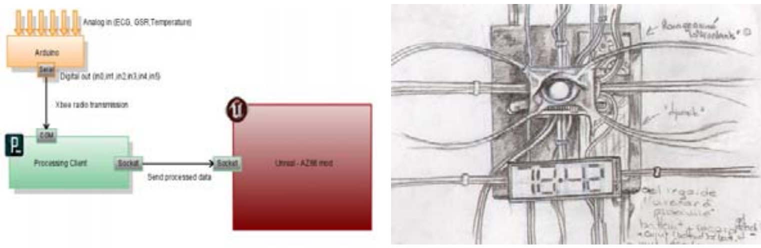


Fig. 1. Hardware and software architecture and artwork

The story brings the player to pass three tests with his physiological states being taken into account. The originality of this game is that the player thinks he always has the control on the game because he has (sometimes fake) bio-feedbacks:

- **Defuse the Bomb:** this situation is expected to be the most stressful since bio feedbacks are fake and not taken into account. The user is powerless.
- **Shoot the Enemy:** the number of enemies is proportional to the player arousal level (GSR) and heart rate.
- **Exit the Maze:** bio feedbacks are taken into account to disturb the player.

At the end of the game, the player gets a global feedback, including his lot in the story and showing how his (real) body reacted during the game, linking the events to the different challenges.

Beyond the rich interaction provided by the dynamic adaptation of game, we expect to design a game validation tool. It will allow the game development team to compare the effective effects of a specific gameplay with the expected ones.

References

1. Picard, R.W.: Affective computing. MIT Press, Cambridge (1997)
2. Sung, M., Pentland, A.: Minimally-Invasive Physiological Sensing for Human-Aware Interfaces. In: HCI International, Las Vegas, Nevada, USA (2005)
3. Loriga, G., Taccini, N., De Rossi, D., Paradiso, R.: Textile Sensing Interfaces for Cardiopulmonary Signs Monitoring. In: IEEE Engineering in Medicine and Biology Conference, Shanghai, China (2005)
4. Weber, W., Glaser, R., Jung, S., Lauterbach, C., Stromberg, G., Sturm, T.: Electronics in Textiles: The Next Stage in Man Machine. In: 2nd CREST Workshop on Advanced Computing and Communicating Techniques for Wearable Information Playing Interaction, Nara, Japan (2003)
5. Lisetti, C., Lerouge, C.: Affective Computing in Tele-home Health. In: 37th Hawaii International Conference on System Sciences (2004)
6. Matt, J.S.: Relax to Win - Treating children with anxiety problems with a biofeedback videogame (2003), `http://medialabeurope.org/research/library/Sharry_Relax_2003.pdf`
7. Magerkurth, C., Cheok, A.D., Mandryk, R.L., Nilsen, T.: Pervasive games: bringing computer entertainment back to the real world. Comput. Entertain. (2005)
8. Dupire, J., Gal, V., Topol, A.: Physiological Player Sensing: New Interaction Devices for Video Games. In: International Conference on Entertainment Computing, Pittsburgh (2008)

WHO AM I? : A Art Ludic Installation in Virtual Reality

Sophie Daste and Karleen Groupierre

ATI (Arts et Technologies de l'Image) Department of Art and
CEN (Création et Editions Numériques) Department of Hypermedia
University of Paris8 at Saint-Denis

Abstract. This paper presents the Art Ludic Installation in Virtual Reality, named "WHO AM I?".

Keywords: Art, Ludic, Installation, Virtual, Augmented, Reality.

1 Introduction

WHO AM I ? is an installation adapting in augmented reality the principle of the game "Le Scotché". Imagine that instead of having a card stuck to your forehead, a virtual head placed on your shoulders reacts emotionally to real time actions…

The simple idea of the change of head has come up after a discussion about the no-limited capacities of the Virtual Reality. So we just thought it was funny to change our face with a virtual head! So The "WHO AM I?" Project is born!

2 Technical Points

We developed "WHO AM I?" with the software Virtools Dassault Système. This software is a platform to compute graphics in real time. We installed the plugging named VirtoolKit, to use a patterns system to enable a tracking functionality within the Virtools environment. We have two computers working in network. This networking is necessary because the software Virtools can't handle two monitors in full screen in the same time.

3 Construction of the Box

We built the box ourselves. We had a specific idea about the construction of our box. We wanted to hide the computers and other technical objects.

So we built a 49x50cm by 88cm high box, suspended 1.50m above the ground. This box is suspended by 4 chains 1.50m long, screwed to the 4 corners of the wooden support board. These 4 chains are joined about 50cm above the box and are linked to a 5th chain which is about 2m long. On two of the opposing faces 35x25cm rectangles are cut to make the two 17inch screens visible. Inside the box, three

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 329–330, 2009.

shelves are installed, the two above supporting the portable computers, the lower one a printer. The customised keyboards are laid in front of the cut faces, on the small shelf below the cut-out area.

4 Design Points

The design interface is very important in this project. It determines the atmosphere of this Art Ludic Installation. We wanted the design of this project to look like something elegant and fun. So like we realized virtual portraits, we have chosen to draw wallpaper to post them.

So we created original wallpaper to define a good atmosphere corresponding to the virtual head.

We also created an atmosphere around this installation, so we have chosen a real wallpaper and we made frames with paper. We created a presentation video where you can see a simulation of a tracking with 2D pictures.

A musician friend composed for this installation three musical pieces with acoustic guitar. This music style with the "WHO AM I?" design breaks with the association between virtual reality and electronic picture and music.

We want really created a special atmosphere. We want created the "WHO AM I?" specific atmosphere.

5 Description of the Artistic Goals

Computer science permits a passage from the real world to a virtual world. Here we transform a real interaction (the old game) to a virtual interactivity (the set-up). In creating this adaptation, particular care must be taken when selecting the platform because each tale, each memory, each dream has a lexical or emotional field that we must make ours to best interpret it. The interpret must therefore judiciously select the medium or platform that will best express his ideas. Here, augmented reality in the form of the installation is the platform that makes the most sense for the re-interpretation of this game.

Thanks to the augmented reality devices, the audience is more included by digital arts. In the "WHO AM I?" device, beyond the gaming experience, we propose to the players a communication through a machine interface. This man-machine interface allows to play with reality by adding, by augmenting the players for the duration of a game.

During this game, the spectator lives a symbiosis with the 3D head, they cannot be dissociated from each other without destroying this hybrid of an augmented spectator. These augmented spectators form an army of machine hybrids: cyborgs.

Affective Interaction: Challenges at the Ubiquitous Computing Times

Stephane Gros, Jerome Dupire, and Stephane Natkin

CNAM - CEDRIC, 292 rue St Martin 75003 Paris, France
stephane.gros@gmail.com, {dupire,natkin}@cnam.fr

Abstract. Due to the diversity of studies towards detection of affective state of a user, it is hard to distinguish amongst them toward specific needs. Our approach consists in realizing a comparative study taking in account specificities such as mobility to be able to choose the best settings for a specific set of constraints. These issues will be applied in the context of the PLUG project which aims at creating an ubiquitous affective game that will take place in a museum.

Keywords: Affective computing, biofeedback, physiology, emotion.

1 Introduction

During the last fifteen years, the theoretical background on human emotion mechanisms has dramatically evolved. Whereas it was formerly considered as a non scientific matter, it is now obvious that **emotions play a significant role in decision making**.

In 1887, James and Lange proposed a theory of emotions suggesting that emotion is largely due to the experience of bodily changes.

In 1929, Cannon and Bard suggest another theory arguing that emotion triggers or stimulates behaviors, as opposed to James-Lange theory.

In Lazarus' theory (**appraisal** theory - 1991), even if some emotions can be triggered unconsciously, emotions arise from the conscious cognitive evaluation of an event. Scherer further developed the appraisal theory of emotions.

Damasio studied patients having decision-making and emotional disorders due to localized brain damage causing both a degradation in the ability to feel and express emotions and decision-making troubles. His observations led him to the hypothesis that emotions play a biasing role in decision making, probably acting a bit like a heuristic: preventing an infinite exploration of rational possibilities and evaluating outcomes *a priori*.

Based on this idea that emotions are needed for rational behavior, Picard defined a field of study for computer science focused on the implication of emotion in Human Computer Interactions (HCI) called Affective Computing.

2 Previous Work

2.1 Affective Computing

Wastell and Newman showed that physiological measures have a valuable role to play in system validation.

S. Natkin and J. Dupire (Eds.): ICEC 5709, pp. 331–332, 2009.

Huber et al. and Ang et al. used prosodic speech analysis to detect anger in utterances respectively with acted data and naturally occurring data.

D'mello et al. built an intelligent tutoring system using affect-sensing technologies (facial expressions, posture information, conversational cues).

Meehan et al. studied presence in stressful virtual environments using physiological measurement (Heart Rate (HR), Galvanic Skin Resistance (GSR) and Skin Temperature).

2.2 Affective Gaming

Iida et al. developed a metric of entertainment for board games only based on the characteristics of the game using variants of chess game.

Gerasimov used physiological measurement as a gameplay mechanism towards personal health monitoring.

Syked and Brown correlated gamepad interactions during play experience with the difficulty of the game. They showed that the gamepad provides physiological information on the user, though further study is needed.

Mandryk et al. used fuzzy logic to extract arousal/valence information from physiological data successfully correlated with subjective emotional evaluations of subjects (e.g. GSR with fun, Electromyogram (EMG) with challenge).

Beaume et al. used game states as a measure of flow in Pac-man game to evaluate fun and difficulty in the game.

3 Researches

We expect that the comparative study of physiological signals acquired in conditions that allow comparison, from normalized stimuli, would permit to determine the set of sensors that best fits specific constraints. From that, we intend to extract a smaller set of physiological measures that are compatible with mobility in terms of intrusivity, encumbrance and dataflow, and focus on noise resistance for these specific biofeedbacks.

An experimental framework for the study of the effect of emotions on physiological measures would also enable the comparison with other measures such as measures on the standard interaction which could complement more intrusive measures.

The evaluation of user reactions to software events opens the possibility of developing new strategies of adaptation and optimization towards user-friendliness. It also allows to consider using physiological data for design validation for software development which is a crucial issue generally limited to subjective report today.

You Are Here

Luc Courchesne

Artist, Professor, Director of the Art and Design Center, University of Montreal

Abstract. The continuing expansion of the visible domain, increasingly including and giving force (and reality) to the creations of our imagination and to formerly inaccessible places and dimensions of our unfolding universe, has produced a radical shift of the shared concepts of time and space. Suddenly, the statement « you are here » has a very different meaning. Thanks to the proliferation and increased mastery of interactive, immersive and socially engaging media, we have passed beyond the simple curiosity for technologies to find ourselves, as MacLuhan professed, dressed and immersed in them. Immersed indeed we are, collectively engaged in a « reality jam » with the physical, the artificial, the augmented, and the virtual. In the process, we have turned from spectators to users, visitors and now inhabitants of our own crafts. Questions such as « Who am I? », « Who are you? », « Where are we? », « How are we supposed to behave? », are now being asked with new relevance and urgency. Survival skills such as conversation, posture, attitude, or simply what to wear, are being reformulated on the fly by the natives and immigrants of this new territory that does not always perfectly accommodates the biological features we inherited from evolution.

The Author. Luc Courchesne took part in the emergence of new media arts in the nineteen eighties when, as a videographer inspired by a generation of experimental filmmakers such as Michael Snow and Hollis Frampton, he was introduced to computers and interactive video technologies. Initially, his work focused on portraiture which he attempted to renovate with his use of interactive technologies to simulate conversion. He later moved toward landscape art which he made interactive and immersive with the Panoscope 360°, a system and method he invented to, in his words, « transform the spectator of the artwork into a visitor ». He created to date over 30 installation works that were exhibited in close to 100 venues the world over including the ICA in Boston in 1983, the Museum of Modern Art in New York in 1994, The ZKM/Karlsruhe since 1997, Paris's La Villette in 1999, Helsinki's Kiasma in 1997 and 2003, The Museum of New South Wales, in 1996 and 2001, the Montreal Museum of Fine Arts in 2007 and the National Art Museum of China in 2008. He was awarded the Grand Prize of The ICC Biennale in Tokyo (1997) and an award of distinction at Prix Ars Electronica in Linz (1999).

Courchesne, who graduated from the Nova Scotia College of Art and Design (1974) and from the Massachusetts Institute of Technology (1984), is now director of the School of industrial design at University of Montreal, a board member of the Conseil des arts et des lettres du Québec and a founding member of the Society for Arts and Technology.

S. Natkin and J. Dupire (Eds.): ICEC 5709, p. 333, 2009.

Game Experience May Vary: Understanding Play

Gonzalo Frasca

Researcher in Game Design, Director of Powerful Robot studio, Uruguay

Abstract. Why do we call it game research and not play research? For the last decade of videogame studies, most of the attention has been paid to games as formal entities. At first, games seem easier to understand: they generally have clear rules and goals. They would be perfect machineries with formal mechanics if it was not for one factor: humans and their stubborn love for misbehaving. This talk provides an ontological approach to play and games and will analyze the relationship between the two concepts by taking into account the player's mindset. If this sounds too theoretical and abstract, there's no need to worry. It holds the key to better understanding the differences between casual and hardcore games.

The Author. Gonzalo Frasca (1972) is a game developer & researcher. He's co-founder of Powerful Robot Games, a game studio based in Uruguay with clients including Cartoon Network, Lucasfilm, Mattel and BBC. His latest hit game for Cartoon Network, AwesomeHouseParty.com, gathered over 22 million player accounts. Frasca obtained his PhD on games at the Center for Computer Games Research in Denmark. His research focuses on game design and rhetoric, particularly educational and political games. In 2003, he co-developed the first official videogame ever created for a US Presidential campaign. In a previous life, he blogged at Ludology.org

S. Natkin and J. Dupire (Eds.): ICEC 5709, p. 334, 2009.

The New Pact: How Online Worlds Forge a New Form of Alliance between Players and Designers

Nicolas Gaume

Founder and CEO of Mimesis Republic

Abstract. Today's Pop Culture is tomorrow's classic. Entertainment is often able to reflect a society's questions and anxieties of the day. We will discuss how and why today's Digital Generation has replaced their grandparents' heroes and their parents' anti-heroes by their very own avatars. How and why the empathy with heroes that has traditionally been utilized by movies has been gradually overtaken by the personalized and immersive experience of games as lived through the eyes of an avatar? How multiplayer activities are now establishing new social paradigms where these avatars help us all deal with the natural schizophrenia of everyday life by allowing us to explore our possible selves. The Social Networks boom has demonstrated our appetite for connections with others; and games our appetite for second lives, our desire to involve ourselves and master complex worlds. Perhaps, tomorrow new public Agoras will be created from a new generation of Virtual Worlds built using the grammar that video games have forged over the last 40 years…

The Author. Nicolas Gaume founded in 1990, at the age of 19, Kalisto Entertainment, a video game development studio. Over its 12 years of life, Kalisto produced more than 50 titles for PC and consoles. Kalisto also worked with Orange to launch their mobile games efforts. The company employed a staff of over 350 people in the USA, Japan and France. Besides his Kalisto venture, Nicolas Gaume founded in 1994, NGM Productions, a children book publishing company in China. NGM published popular French series such as "Père Castor" (Flammarion) or "Les Incollables/Brain Quest" (Play Bac) between 1994 and 1999, when the company was sold. He also co-founded, in 1995, one of the first French web agency, Wcube, sold to Swedish group, Framfab, and, in 2000, a wine promotion site, winealley.com. He was also board member of the French media group Sud-Ouest, of the TV animation production company, Xilam & of the German data management software publisher, NXN, sold in 2004 to Avid. Between July 2002 and May 2005, Nicolas has worked as a consultant advising media group such as AOL, Orange or Lagardère and game companies such as Codemasters or Ubisoft, primarily on their development & partnerships strategies. For thee years, Nicolas Gaume was then Senior VP & GM of the mobile games & applications division of Cellfish Media, a Lagardère company. Early 2008, he founded a new company, Mimesis Republic, focusing on the design, development and operation of virtual worlds and social networks. He also currently serves as chairman of the French national video game development school –ENJMIN- and president of the French game developers association, SNJV.

S. Natkin and J. Dupire (Eds.): ICEC 5709, p. 335, 2009.

The International Game Developer Association (IGDA) Education Special Interest Group (EdSIG)

Susan Gold

Professor, Full Sail University; Executive Director, The Global Game Jam™, Director of Game Program Review, Chairperson, IGDA Education SIG

Abstract. Developed in 2006 the IGDA EdSIG mission is to create community resources that will strengthen the academic membership of the IGDA while enhancing the education of future and current game developers. Since the inception of the SIG there have been two major projects: 1) 2008 IGDA Curriculum Framework which has become the standard for game education; 2) The Global Game Jam™ where experimentation, innovation and creativity are the main components of it's success. The Global Game Jam™ creates collaborations and partnerships globally to promote game studies.

Author. Susan Gold recently joined the faculty of Full Sail University's Graduate Program in Game Design. Prof. Gold's personal focus is on collaboration, which resulted in the now annual Global Game Jam™ (http:// globalgamejam.org) which she found and serves as the Executive Director. The Global Game Jam™ (GGJ) was established in 2009 with great success and media attention, 1650 participants creating 370 video games. GGJ brings together talented individuals and teams from around the globe and rallies them around a central theme, for which they have 48 hours to create their game. Susan serves as chairperson of the International Game Developers Association (IGDA) Education Special Interest Group (EdSIG) and her work with the IGDA EdSIG has created numerous resources for instructors in game education. Susan spearheaded the industry and academic collaborative project, the 2008 IGDA Curriculum Framework. Currently, her EdSIG work is focused on building an outreach campaign to meet the needs of educators worldwide, while continuing to develop tools & resources for educator professional development. Ms. Gold is a partner in the Game Program Review, a consulting firm specializing in game education game curricula. Susan is an artist, teacher, and activist with a specialization in digital art, new media and videogames. Her artwork and writing has been featured in numerous galleries, museums and publications.

S. Natkin and J. Dupire (Eds.): ICEC 5709, p. 336, 2009.

Story of a Video Game Workshop: "Ico", an Interactive Fairy Tale for Children Less Interaction

Michael Stora

Clinical Psychologist, Co-creator of the “Observatory of Digital World in Human Studies”, author of “Guérir par le jeu par les jeux vidéo, une nouvelle approche thérapeutique” Collection les Presses de la renaissance, Paris, 2005

Abstract. The Since several years, I use the video game as a therapeutic object. This new mediation seems pertinent for children that we qualified "limit". Those children are more in the game in the gambling sense, than in the Winnicot's play. The video game is relevant in a clinic sense for two reasons essential. The first reason is that mediation is in picture and children and adolescents have an intimate attachment to these who have the power to do emerge affects and words that in generally are repressed. The second reason is this meeting between narcissism pathology and the use of video game as a place of virtue's passage in act. The acting becomes symbolic and the conflict showed by the narration of the video game is in a dynamic point between drive and narcissism.

The Author. After my education as a film-maker, I became a clinical psychologist, and, ever since I am constantly preoccupied by the unconscious links between human being and images. In 2000, I created a Partnership Association named: "Observatory of Digital World in Human Studies", for developing new theories on Interactivity, this new connection to images. I, moreover, used video games as a therapeutic tool for children with violent behaviour troubles. In 2008, I launched the Company "Implicit Game" with the aim of creating video games for healing. I am currently working on a project of an on-line game for youngs adutls and teen agers suffering of cancer.

S. Natkin and J. Dupire (Eds.): ICEC 5709, p. 337, 2009.

Japan's Arcade Games and Their Technology

Y. Sambe

CTO Executive Director of TAITO Corp.

Abstract. The Japanese computer entertainment market is $20 billion in size. Of that, the arcade market makes up the biggest share of revenue ($6 billion), with the home market and mobile phone market following at $3.5 billion and $2 billion, respectively. Abroad, however, home console gaming dominates and revenues from arcades make up only a small portion of the total market. In this session, I will introduce the arcade machines and supporting technology, discuss reasons for the enduring widespread popularity of arcade gaming in Japan, and explore potential directions for the future of arcade technology .

The Author. 1979 Joined Taito Corporation. Worked in the field of arcade machine research and development. 1992 Invented network karaoke (via server & client karaoke machines), transforming the service into a viable business. 1995 Established the home network karaoke business. 1995 Promoted to Executive Director. 1999 Founded TAITO's mobile content business. 2004 Assigned as Senior Executive Director. 2006 Created TAITO's Key Technology Research Center R&D Division for new business 2007 Named TAITO's Chief Technology Officer (after resigning as Senior Executive Director).

Awards. 1993: Nihon Keizai Shimbun: Recognition for establishing the network karaoke business. 1995: Nihon Keizai Shimbun: Recognition for contributions to the home network karaoke business. Member of The Institution of Professional Engineers, Japan Technology Committee Member of CESA (Computer Entertainment Supplier's Association)

S. Natkin and J. Dupire (Eds.): ICEC 5709, p. 338, 2009.

Erratum to: Entertainment Computing – ICEC 2009

Stéphane Natkin and Jérôme Dupire

Conservatoire National des Arts et Métiers, CEDRIC, 292, rue St. Martin,
75141 Paris Cedex 03, France
{stephane.natkin,jerome.dupire}@cnam.fr

Erratum to:
S. Natkin and J. Dupire (Eds.)
Entertainment Computing – ICEC 2009
DOI: 10.1007/978-3-642-04052-8

The book was inadvertently published with an incorrect name of the copyright holder. The name of the copyright holder for this book is: © IFIP International Federation for Information Processing. The book has been updated with the changes.

The updated original online version for this book can be found at
DOI: 10.1007/978-3-642-04052-8

S. Natkin and J. Dupire (Eds.): ICEC 5709, p. E1, 2009.

Author Index