LNCS 5309

Scott M. Stevens
Shirley J. Saldamarco (Eds.)

Entertainment Computing – ICEC 2008

7th International Conference
Pittsburgh, PA, USA, September 2008
Proceedings

ifip

Springer

Lecture Notes in Computer Science 5309

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

Scott M. Stevens Shirley J. Saldamarco (Eds.)

Entertainment Computing – ICEC 2008

7th International Conference
Pittsburgh, PA, USA, September 25-27, 2008
Proceedings

Springer

Volume Editors

Scott M. Stevens
Carnegie Mellon University
Entertainment Technology Center and Human Computer Interaction Institute
Pittsburgh, PA 15213, USA
E-mail: sms@cs.cmu.edu

Shirley J. Saldamarco
Carnegie Mellon University
Entertainment Technology Center
Pittsburgh, PA 15213, USA
E-mail: shirley@cs.cmu.edu

Library of Congress Control Number: 2008938141

CR Subject Classification (1998): H.5, H.4, H.3, I.2, I.3, I.7, J.5

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

ISSN	0302-9743
ISBN-10	3-540-89221-4 Springer Berlin Heidelberg New York
ISBN-13	978-3-540-89221-2 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12567492 06/3180 5 4 3 2 1 0

Preface

The 7th International Conference on Entertainment Computing, under the auspices of the International Federation for Information Processing (IFIP), was held September 25–27, 2008 in Pittsburgh, Pennsylvania. Based on the very successful first international workshop (IWEC 2002) and the following international conferences (ICEC 2003 through ICEC 2007), ICEC 2008 was an international forum for the exchange of experience and knowledge amongst researchers and developers in the field of entertainment computing. ICEC is the longest established and most prestigious conference in the field of entertainment computing.

The conference provides an interdisciplinary forum for advanced research in entertainment computing, broadly defined. ICEC is truly international with leading experts from 14 nations representing academia and industry attending this year's conference. These leaders presented their newest research, insights, products and demonstrations.

Although the field of entertainment computing is thought of as new, in fact modern digital computer games go back over 45 years with games such as Spacewar developed in 1961. This is not to say entertainment computing is limited to computer games. As evidenced by papers in this volume, entertainment computing covers virtually every aspect of today's recreational diversions. With evocative titles like: "Development and Evaluation of a Centaur Robot," "Analysis of Japanese Folktales for the Purpose of Story Generation," "Fear Inducer: A Mixed Reality Audio Experience," "Interactive Multimedia Contents in the IllusionHole," "Robust Interactive Storytelling for Automatic TV Content/Story Production," "Designing Toys That Come Alive: Curious Robots for Creative Play," "Musical B-boying: A Wearable Musical Instrument by Dancing," "Hybrid Visual Tracking for Augmented Books," "Hitch Haiku: An Interactive Supporting System for Composing Haiku Poem," and "Experiences Employing Novice Wizard Operators in a Gallery Setting," ICEC 2008 papers spanned entertainment experiences from robots to television, music to audio enhancing emotion, and dancing to poetry.

Of course, traditional game development was well represented with papers like: "Immersion, The Greatest Hook," "Agents That Relate: Improving the Social Believability of Non-Player Characters in Role-Playing Games," "Game Bot Detection Based on Avatar Trajectory," "Creating an Emotionally Adaptive Game," "Towards Emotional Characters in Computer Games," "Newsgames: Theory and Design," and "Physiological Player Sensing: New Interaction Devices for Video Games."

The host institution for ICEC 2008 was the Entertainment Technology Center of Carnegie Mellon University. Few institutions in the world embody the interdisciplinary characteristics underlying entertainment computing as does the ETC. Founded by Don Marinelli, Professor of Drama and Arts Management, and Randy Pausch, Professor of Computer Science and Human-Computer Interaction, it is a marriage of the worlds of computer science and entertainment at the most fundamental level; a joint partnership of the College of Fine Arts and the School of Computer Science. Both CMU entities are amongst the nation's top institutions in their fields. ETC's fusion of disciplines is a testament to Carnegie Mellon's longstanding support of interdisciplinary research, where first-class faculty and students perform cutting-edge, world-changing research; research that is not limited by the bounds of departments or fields. Likewise, ICEC 2008 brought the best of interdisciplinary work in computing in support of entertainment. We hope the works that follow give the reader a sense of the excitement and intellectual atmosphere experienced at this year's conference.

September 2008

Scott M. Stevens
Shirley Saldamarco

Organization

General Conference Chair

Don Marinelli	Carnegie Mellon University, USA

International Program Committee Co-chairs

Sara Diamond	Ontario College of Art and Design, Canada
Sidney Fels	University of British Columbia, Canada
Scott Fisher	University of Southern California, USA
Barnabás Takács	Hungarian Academy of Sciences, Hungary
Hyun S. Yang	Korean Advanced Institute of Science and Technology, Korea

International Program Committee Secretary

Scott Stevens	Carnegie Mellon University, USA

International Steering Committee

Marc Cavazza	University of Teesside, UK
Tak Kamae	Tokyo Women's Medical University, Japan
Lizhuang Ma	Shanghai Jiao Tong University, China
Stéphane Natkin	CNAM, France
Ryohei Nakatsu	Kwansei Gakuin University, Japan
Matthias Rauterberg	TU Eindhoven, The Netherlands
Peter Robinson	University of Cambridge, UK
Andy Sloane	University of Wolverhampton, UK
Bill Swartout	University of Southern California, USA

Organization Co-chairs

Shirley Saldamarco	Carnegie Mellon University, USA
Ben Salem	Eindhoven University of Technology, The Netherlands

Sponsoring Institutions

International Federation for Information Processing
Entertainment Technology Center, Carnegie Mellon University
WMS Gaming

Table of Contents

Posters

How I Learned to Love the Bomb: *Defcon* and the Ethics of Computer Games

Miguel Sicart

Assistant Professor
Center for Computer Game Research
IT University of Copenhagen
http://www.itu.dk
Junior Research Associate
IEG - Oxford University
http://web.comlab.ox.ac.uk/oucl/research/areas/ieg/
GPI - University of Hertfordshire
http://philosophyofinformation.net/centre/gpi/index.html
miguel@itu.dk
www.miguelsicart.net

Abstract. In this paper I present an analysis of the ontology and ethics of computer games from an Information Ethics perspective. This analysis uses the concepts of Level of Abstraction and Gradient of Abstraction, as defined by Luciano Floridi's Information Ethics, applied to the specific study of computer games. The goal of this paper is to argue for the consideration of games as interesting ethical objects and experiences. Computer games appeal to a player capable of ethical reasoning in her interaction with simulated environments and rule systems. This paper provides a theoretical model for the study of the ethics of computer games both as designed objects, and as player experiences.

Keywords: Information Ethics, Computer Game Ethics, Method of Abstraction.

1 Introduction

Computer games are the dominant paradigm of 21st Century entertainment. Their worldwide economic success inspires the technological research that extends the boundaries of game-related computation. But computer games have too a reputation of being intrinsically unethical forms of entertainment that corrupt the values of society by promoting the wrong personal and social virtues.[1] This reputation is caused and exemplified by the simulation of violence that many computer games use as a vehicle for the expression of conflict.[2] There is an ethical tension between the perception of computer games as infantile entertainment, and the demands to the game industry from their core adult audience.

In any case of moral concerns it is necessary to question the ethical origins of this perception. This paper argues for an interpretation of the ethics of games as designed

1 Carnagey et al. (2006).
2 Juul (2005), pp. 92 - 116, and Salen and Zimmerman (2004), pp. 74-78 & 254-255.

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 1–10, 2008.

software systems that are experienced by moral agents. The goal of this paper is to present a model for the ethical analysis of games. This model will use concepts from Luciano Floridi's Information Ethics[3], applied to the analysis of the independent multiplayer computer game *Defcon* (Introversion Software: 2006).

This is a paper on Information Ethics. As such, it applies the theoretical framework proposed by Floridi to the analysis of computer game. The terminology, and the Method of Abstraction are direct appropriations from Information Ethics, but their application to computer games is entirely original. Furthermore, concepts like Procedural and Semantic Gradients of Abstraction, are original additions to Information Ethics theory, first presented in this paper.

Information Ethics provides a comprehensive framework for the analysis of computer games for three fundamental reasons: first, the focus on information as an ontology facilitates the analysis of computer games as systems, beyond their audiovisual and aural components, yet integrating these in the whole meaning of the game experience; second, Information Ethics provides a strong vocabulary, with roots in Information Theory and Computer Science, which affords precise definitions of systems, and their interaction modes, from an ethical perspective; third, Information Ethics is a constructivist theory with a strong anthropological model[4], which in turn can be used to outline a model of ethical player. In any case, this ethical player model is beyond the scope of this paper, and will only be suggested when presenting the informational model of computer games as ethical experiences.

This paper is not an answer to effect studies research, nor it intends to clarify the possible ethical problems of violent games and children.[5] This is a philosophy paper, and as such it contributes to the understanding of the ethics of computer games by means of an abstract model for the analysis of the moral relations between a player, and a designed system. The model illustrates how the values of the game system and the values of the player are deeply intertwined. These achievements can be used for further analysis of computer games ethics, both from a design, and a psychological or phenomenological theory.

This model will suggest a philosophical argument for understanding the ethical liability of computer games. While it is true that many games simulate violence, basing all their interactions in a glorification of gore, we need to understand their ontology, in order to qualify their ethical nature. In this paper I present an analytical tool that can be used to argue for games as generators of interesting ethical experiences – doing so by critically involving the player in the interaction system. In other words, many computer games require an ethical agent in order to create any type of ethical experience.

2 Computer Games as Ethical Challenges

Defcon can be a game of patience. *Defcon* gives the player command of a nuclear country. The goal is to eliminate as many enemy units as possible, while minimizing loses. In fact, to win the game, players have to lose the least. *Defcon* is a political simulator of atomic warfare, with a very clear message: in nuclear war, the winning condition is a losing condition.

---

[3] Floridi (2003a), (2003b).

[4] Floridi and Sanders (2003).

[5] Funk et al. (2004).

Defcon is a multiplayer computer game played over the Internet. Players are given a limited number of units and resources that have to be distributed in the map. This first stage is a preparation for the nuclear showdown that inevitably will take place at the end of the round, when Defcon One is reached and players can use nuclear warheads. When the game reaches its end state, players will be rewarded with points according to the number of enemy units and civilians eliminated. The winner of the round is the player who loses the least, which implies that targeting large cities is the best strategy.

In the classic western theory of games, as exemplified by Huizinga[6] and Caillois [7], games are perceived as something "separate", as an "unproductive" activity that takes place within boundaries set by rules agreed upon by players[8]. Games are then arbitrary systems that establish constraints. Players have to accept these constraints in order to achieve the goals also proposed by the game system. Finally, games are considered entertainment, pastimes, vehicles for leisure with clear rules and unambiguous outcomes.

In more formal terms, games can be defined as activities in which practitioners interact with a system designed to create obstacles to the achievement of predetermined goals. To play a game is to give supreme but temporary importance to these constraints. For the player, nothing is more important than the rules and the goals of the game. Play is the focus of the self on artificial rules, worlds and goals. This focus brings forth ethical concerns, specially in the case of games played with computers.

In computer games, the rules are embodied in a virtual world. In many cases, the rules and the virtual world can be perceived by external observers as ethically harmless: there is no moral risk in *New Super Mario Bros*[9] because the "violence" is cartoonish. But games like *Defcon*, or *Doom,*[10] or *Counter-Strike*[11] seem to create immediate ethical concern. These games could be understood, in a classic Virtue Ethics approach[12], as tools for the practice of the wrong virtues: violence, killing, inadequate conflict resolution. In this line of thought, a game has to be unethical because the agent is presented with rewards for actions of (simulated) evil, while feedback on the consequences is not provided[13]. The act of "killing" an opponent in a computer game becomes an action with no other consequence than what the game rules determine, therefore alienating the player from her reflections on the consequences of her actions.

This analysis of the ethics of games would then state that if players are faced with simulated evil and rewarded for their non virtuous actions in that gameworld, then players will internalize the lack of consequence for these actions. This process would be the first step in the desensitization effects that some researchers[14] and media believe is the playing computer games.

In essence, this interpretation assumes that there is no difference between the act of launching an atomic bomb over Hiroshima and launching an atomic bomb over

---

6 See Huizinga (1950).
7 See Caillois (1958).
8 For a critical summary of the western understanding of the ontology of games, see Juul (2004), pp. 23-55.
9 Nintendo: 2006.
10 id Software: 1993.
11 Valve Software: 1999.
12 As defined by Aristotle's Nichomachean Ethics. See also Feezeel (2004).
13 Nevertheless, it is necessary to say that the consequences of actions in games do have a system for feedback, embedded in the rule system, and usually tied to the winning conditions.
14 See Carnagey et al. (2006), Funk et al. (2004).

Tokyo in *Defcon*. Thus, playing computer games is an unethical act. This is both a limited understanding of computer games as cultural objects, and a very poor consideration of the ethical capacities of players as moral agents. Still, to counter argue, it is necessary to define what computer games are, and how players experience them. The rest of the paper will transform the alleged ethical shortcomings of computer games into analytical tools for defining computer games as ethical experiences. Computer games are not detached, encapsulated systems of meaning that classic game research has argued for. Computer games have a strong presence in the configuration of our ethical and cultural being, and as such we must describe them.

3 Computer Games as Informational Systems

When describing a game to someone that has never played it, the first stage is to describe the rules and the game mechanics. Players need to know what is allowed and possible, and what is not. Once players have a basic understanding of the rules, play is initiated. The initial states of play are approximations both to the rules and to the environment where we play. This process tries to find the winning strategies while staying true to the rules. Mastering a game is understanding the rules and mechanics of the game, how they interact with each other. Mastery leads to the behavioral patterns with which we play.

From a philosophical perspective, a game can be defined as an informational system: a construction of rules that determine which actions are meaningful or not within a certain experience, and how those actions can be performed. This ontology is based on the theory of Information Ethics. For Information Ethics, "the moral action itself can now be modeled as an information process, i.e., a series of messages (M), invoked by *a*, that brings about a transformation of states *directly* (…) affecting *p*, which may variously respond to M with changes and/or other messages, depending on how M is interpreted by *p*'s methods"[15]. I will adapt this approach to encompass all kinds of *ludic* actions, defined as any action taken by an agent within a game system that is evaluated by a game rule. For example, placing the units in *Defcon*, firing a missile, or using the chat interface to communicate with other players, they all are ludic actions.

In order to understand the ethics of computer games from an informational perspective, it is relevant to define computer games within the terminology of Information Ethics. A computer game, then, is an infosphere, a "context constituted by the whole system of information objects, including all agents and patients, messages, their attributes and mutual relations"[16]. In more classic computer game terms, a (game) infosphere is constituted by all game elements: players and AI agents, environments and gameworld, rules and game mechanics, and the interaction modes in the space of possibility[17]. In computer games there are agents, human or controlled by the game software. From a certain perspective, all agents operate similarly: interacting with the game state[18] via game mechanics constrained by game rules. Playing a game is an act of agency within an infosphere, understanding agency as the interaction by means of exchange of information with a system

[15] Floridi (2003a).
[16] Floridi (2003), p. 8.
[17] Salen & Zimmerman (2004), pp. 66-67.
[18] See Juul (2005), chapter 3.

and other agents.[19] This exchange of information is done via the methods[20] of the different game objects, as constrained by the game rules.

In *Defcon,* players construct a strategy by carefully placing units in the places they are afforded by the rules, relative to their initial state in the game. Players manipulate the gameworld and their relations to other players in ways sanctioned and determined by the rules. Agency in *Defcon*, as in any computer game,is limited by the game rules and the mechanics afforded to players. The relation between players and the game environment will take place within the boundaries of those mechanics. The game as infosphere is determined by the design of the game.

From an ethical perspective, it is crucial to establish that the infosphere was designed to afford agency and allow for the flow of information in specific modes. Any system designed to modify or enhance agency, any system that is *scripted*[21]for creating behaviors, has to raise ethical awareness. Altering agency can have ethical implications in the ontological status of the agents.

For instance, *Defcon* is designed to encourage conflict among players. All the mechanics and rules of the infosphere are geared towards creating conflict. Players of *Defcon* cannot find negotiated outcomes - the goal of the game is to simulate nuclear war and to reward those that annihilate more units of the opponent. The design of the game, as the design of any infosphere, is loaded with ethical values set to be experienced. More specifically, *Defcon* rewards those players who lose the least. Conventionally, wargames reward players who take out other players, regardless of their losses. *Defcon* modifies that rule, which leads to an ethical impact on the player experience.

The morals of game agents are not determined by the ethical values of the game design. Any agent in an infosphere, and particularly any human agent, has to be considered a moral agent. A moral agent is capable of ethically relating to the whole system, reflecting on her own values and ethics. Furthermore, moral agents are able of acting upon these values, modifying the actual ethics of the infosphere as experienced. While an infosphere can have ethical values imprinted in its design, it is the actual interaction of a moral agent with those ethical affordances and constraints[22] what constitutes the ethics of a given infosphere. The informational ethics of a game system can only be analyzed when the game is experienced by agents, and not exclusively by its design, or its appearance.

Agents exert their ethical capacities by what Floridi has defined as creative stewardship. Human agents, from an Information Ethics perspective, ought to be considered *homo poieticus*, since they concentrate “not merely on the final result, but on the dynamic, ongoing process through which the result is achieved”[23]. Agents within an infosphere are not only in charge of exchanging information, but also of preserving the nature of the system, producing meaningful interactions. In computer games, this translates to the ethical responsibility for playing without cheating, not allowing other players to grief or harass the community, or developing and sharing interesting strategies for beating the game. Players

[19] See Floridi and Sanders (2004b) for a more detailed Information Ethics approach to the question of agency.

[20] Methods should be understood in the Object Oriented Programming paradigm, as the mechanisms an object has to access and/or manipulate data within other objects. See Weisfeld (2000).

[21] See Latour (1992).

[22] See Norman (2002).

[23] See Floridi and Sanders (2005).

as ethical agents are not mere input providers: they understand the nature of their actions within the semantics of the infosphere, and they act ethically.

This moral understanding in games is determined by two elements: the player as agent, and the cultural being that experiences play[24]. As players, we construct our agent values with those ethical affordances and constraints provided by the system. In *Defcon,* players' ethics reckon how the system encourages treason, since there is only one winner at the end of the game, but alliances with other players are encouraged. From an information ethics perspective, the player has a number of methods that allow her informational interaction with the other agents of the system. Those methods simulate nuclear war with the goal of devastating opponents and provoking casualties in their cities. The system is designed to encourage and reward that agency.

Players are not only input providers within the game system - their ethical configuration is also dependent on the ethics of the agent that becomes a player. The ethical agent outside of the gameworld is also relevant for understanding the ethics of computer games. *Defcon* is not only a state machine that simulates the conditions of conflict: it simulates nuclear war and its outcomes, representing them by means of an aural output system. Both the aural system and the simulation of nuclear war, what I will define as the semantic layer of the game, are interpreted by the human agents using their cultural and ethical knowledge.

For an analytical perspective, then, the infosphere comprises both the agents that interact with predetermined methods with the state machine, and the moral agents that evaluate the cultural and ethical relevance of in-game agency.

4 The Method of Abstraction and the Ethics of Computer Games

The Method of Abstraction[25] provides a framework for analyzing the relation between agents and designed systems. In this paper I will use the concepts of Level of Abstraction (LoA henceforth) and Gradient of Abstraction (GoA henceforth). A Level of Abstraction is " finite but non-empty set of observables. No order is assigned to the observables"[26]. A Gradient of Abstraction "is a formalism defined to facilitate discussion of discrete systems over a range of LoAs. Whilst a LoA formalises the scope or granularity of a single model, a GoA provides a way of varying the LoA in order to make observations at differing levels of abstraction"[27]. These two concepts, applied to computer games, allow the formulation of a model for the analysis of the ethics of computer games.

In computer games there are two dominant GoAs: one is limited to the direct interaction between agents and the state machine by means of game mechanics, the other comprises the game system as simulation and agents as ethical agents.

This first GoA defines all the input/output operations performed by and for the modification of the game state within the limitations of the rule system. For example, the action of selecting a unit in *Defcon* invokes a LoA in which the user interface, the class and particular instance of the unit, and the response from the state machine are relevant. I will define this GoA as the syntatic or procedural GoA. The syntatic/procedural GoA comprises the LoAs that regulate the input/output processes between agents and the

[24] See Gadamer (1975).
[25] Floridi and Sanders (2004a).
[26] Floridi and Sanders: 2004a, p. 10.
[27] ibid, p. 12.

state machine. The syntactic/procedural GoA constitutes the inner mechanisms of the game, it's procedural nature[28] as a system.

The second GoA comprises the game system as simulation and agents as ethical agents. This expands the previous GoA by adding a semantic layer. The game system is in this layer more than a simple state machine: it comprises all the aural/aesthetic levels, giving cultural meaning to the elements of the game system. In *Defcon*, the state machine simulates nuclear war. The system behaviors, the semantic levels, its procedural and aesthetic content, they are all designed to be interpreted as war commanded from a nuclear bunker. This GoA comprises the procedural nature of the game with the cultural/aesthetic layers that wrap it, and it can be defined as a semantic/simulational GoA.

The agent in the semantic/simulational GoA is capable of reflecting morally not only about the appropriateness of her actions providing input, but also about the meaning of those actions. This meaning is interpreted both within the perspective of the cultural meaning of the simulation and with her own cultural values[29].

The semantic GoA comprises the procedural GoA, and all the different LoAs present in a computer game can then be visualized within those relations. This conceptualization can be modeled in the next page.

At the heart of the ethical concerns with computer games lies the incapacity of understanding how players navigate in the infosphere. Common accusations against computer games understands them only within the procedural GoA, and thus from the perspective of an agent not concerned with anything else than providing the right input to modify the state machine. Furthermore, there seems to be a common issue to map the simulation within the procedural GoA, which is ontologically incorrect: the mechanical agent is not concerned with the simulation, it's beyond its LoA. When agents interact with the simulation, they use too their ethical capacities, since those are a part of their cultural resources required to understand how and why to play a game.

A common ethical concern raised by games is concerned with desensitization, which in this philosophical context I define as crisis of the agent's ethical tools for evaluating their conduct.[30] This concern is based on the assumption that players allegedly provide input to the system without thinking about the consequences of their actions in the simulation, and thus becoming alienated from the causality reflection.

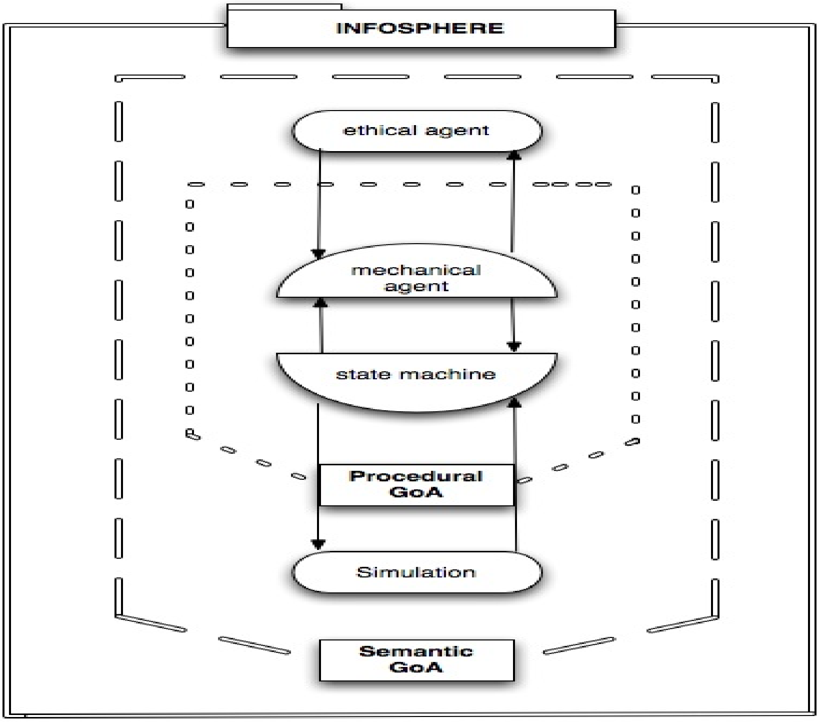
But given this Information Ethics analysis, this concern does not hold true. Any agent in a game operates within two distinct GoAs - one that cares about the procedural elements of the game, from user interface to score system, and another one that encapsulates that procedural part of the game into a larger GoA. In this larger GoA, the semantics of the game are extremely relevant, as they guide player interactions with the system.

---

[28] See Bogost (2006).

[29] This seems to discard any agent that is not human, but that is not the case: there are some LoAs in that GoA that can apply to agents that are not human: every LoA that does not require attention to those values external to the simulation can apply to artificial agents. A more detailed approach to artificial agency in infospheres can be found here: See Floridi and Sanders (2004b).

[30] As opposed to the psychological approach, concerned with the reduction of emotional response: see Funk et al. (2004). A possible continuation of this paper would lead to comparative work in social psychology and philosophy, on the ethical and emotional effects of these games in different ethical agents, mapped to the informational structure of the game, as modeled in this paper.



This process can be understood with an analogy: piano players often need to "warm" their fingers before playing. They do so by playing on the piano a number of exercises. Piano players are not concerned with the semantic elements of the music piece, only with the mechanical interaction with the piano. Once they start playing a piece, though, the semantic layer is also present, so tone, emotion, phrasing, and the aesthetic qualities of playing piano require attention.

This informational perspective on the ontology of games provides an framework for the analysis of the ethics of computer games. It also gives arguments as to why some computer games are not relevant for ethical theory. Ethically irrelevant games are those in which the procedural dominates over the semantic. For instance, in *Tetris*[31]players don't need to understand the simulation in order to successfully interact with the system. To play *Tetris* is enough with an understanding of the rules and mechanics. Abstract games[32], then, are those games that privilege the importance of the procedural over the semantic GoA.

Summarizing, the procedural GoA comprises the design and implementation of the game as a state machine, with the basic mechanics and rules that determine the interaction of input agents with the system. The semantic GoA comprises the layers of meaning that we understand as the gameworld - the reasons why players are emotionally attached to the game, understand how to play it, and take choices. Agents in this gradient are concerned with the player community, the cultural and ethical values of

[31] Alexei Pajitnov, 1985.
[32] See Juul (2005), pp. 130-131.

the game, and the connections of the game with larger infospheres external to the game experience.

The games that taunt players with ethical decision-making, like *Fable*[33] or *Knights of the Old Republic*[34] are flawed because their alleged ethical thinking is placed in the procedural layer: "evil" is not understood as a dominant semantic condition but a procedural one, as a state in the machine. Ethical agents do not require to use their ethical reasoning within the semantic layer in order to take a choice: it is enough to understand the arbitrary ethics assigned to a particular game state, and let the game system evaluate the behavior. These games generate a process of desensitizing the agent of their ethical thinking about the simulation. Since ethics are limited to the procedural layer, agents are deprived of their ethical capacities in favor of an external system that will evaluate their choices.

5 Conclusions

This paper has presented an analytical framework of the ethics of computer games as informational environments. This framework is based on Information Ethics, and provides a comprehensive approach to the technological and cultural challenges of computer games' ethics.

The model I have presented in this paper can be used not only to analyze the ontology of games as ethical systems, but also to provide a philosophical answer to issues concerning games and unethical content. That argument goes as follows: any computer game requires a player who understands not only the basic rules and interactions, but also the meaning of the game world, of the simulation. A player that can understand that sematic layer is an ethical agent, and as such will always use moral maturity to interpret and experience the game. In other words – players are ethical agents that require ethical thinking to interact with the game.

A subject for further research is the meaning of ethical thinking for players. I would argue that ethics, like Aristotle wrote, is a practical science that requires maturity, a constant practice towards perfection. Similarly, the ethical maturity of players evolves with time and experience of games, and thus social instances like age regulation codes, are effective tools for guaranteeing the successful development of players as ethical agents.

Computer games offer experiences of seemingly impossible worlds that put players in the center of epic tales of heroism. But the moral dimension of games is still in its infancy. With this paper I have introduced a framework for understanding games as ethical systems. The next step is to put into practice the lessons learnt from this reflection, since the possibility of creating engaging ethical gameplay is the true promise of any next generation gaming.

References: Literature

Bateman, C., Boon, R.: 21st Century Game Design. Charles River Media, Massachusetts (2005)

Bogost, I.: Persuasive Games. The Expressive Power of Videogames. The MIT Press, Massachusetts (2006)

[33] Lionhead: 2004.

[34] Bioware: 2003.

Caillois, R.: Man, Play and Games. Translated by Meyer Barash. University of Illinois Press, Urbana (1958/2001)
Carnagey, N.L., Anderson, C.A., Bushman, B.J.: The effect of video game violence on physiological desensitization to real-life violence. Journal of Experimental Psychology (2006)
Feezell, R.: Sport, play, and ethical reflection. University of Illinois Press, Urbana (2004)
Floridi, L., Sanders, J.: Internet Ethics: the Constructionist Values of Homo Poieticus. In: Cavalier, R. (ed.) The Impact of the Internet on Our Moral Lives, New York, SUNY (2005)
Floridi, L., Sanders, J.: Levellism and the Method of Abstraction, Information Ethics Group Research Report (2004a), http://www.philosophyofinformation.net/pdf/latmoa.pdf
Floridi, L., Sanders, J.: On the Morality of Artificial Agents. Minds and Machines 14(3), 349–379 (2004b)
Floridi, L.: On the Intrinsic Value of Information Objects and the Infosphere. Ethics and Information Technology 4(4), 287–304 (2003a)
Floridi, L.: Two Approaches to the Philosophy of Information. Minds and Machines 13(4), 459–469 (2003b)
Funk, J.B., Baldacci, B., Heidi, Pasold, Tracie, Baumgardner, Jennifer: Violence exposure in real-life, video games, television, movies, and the internet: is there desensitization? Journal of Adolescence 27, 23–29 (2004)
Gadamer, H.G.: Truth and Method (1975); Translated by Weinsheimer, J., Marshall, D.G. Continuum, New York (2004)
Huizinga, J.: Homo Ludens. The Beacon Press, Boston (1950)
Juul, J.: A Certain Level of Abstraction. In: Proceedings of the III DiGRA Conference (2007)
Juul, J.: Half-Real. Video Games between Real Rules and Fictional Worlds. The MIT Press, Massachusetts (2005)
Latour, B.: Where are the Missing Masses? - the Sociology of a Few Mundane Artifacts. In: Bijker, W., Law, J. (eds.) Shaping Technology/Building Society, pp. 225–258. The MIT Press, Massachusetts
Norman, D.: The Design of Everyday Things. Perseus, New York (2002)
Salen, Katie, Zimmerman, Eric: Rules of Play. Game Design Fundamentals. The MIT Press, Massachusetts (2004)
Suttton-Smith, B.: The Ambiguity of Play. Harvard University Press, Cambridge (1997)
Weisfeld, M.: The Object Oriented Thought Process. Sams Publishing, Indianapolis (2000)

References: Games

Bioware (2003), Knights of the Old Republic
Blizzard Entertainment (2005), World of Warcraft
CCP Games (2003), Eve Online
id Software (1993), Doom
Io Storm (2000), Deus Ex
Lionhead Studios (2004), Fable
Newsgaming.com (2004), September 12th
Nintendo (2006), New Super Mario Bros
Pajitnov, A. (1985), Tetris
Quantic Dream (2005), Fahrenheit
Rockstar Games (2003), Manhunt
Sony (2005), Shadow of the Colossus
Valve Software (1999), Counter-Strike

When Items Become Victims: Brand Memory in Violent and Nonviolent Games

André Melzer, Brad J. Bushman, and Ulrich G. Hofmann

A. Melzer, Université du Luxembourg, Luxembourg
andre.melzer@uni.lu
B. J. Bushman, University of Michigan, Ann Arbor, USA, and VY University Amsterdam, The Netherlands
bbushman@umich.edu
U. G. Hofmann, University of Luebeck, Germany
hofmann@imis.uni-luebeck.de

Abstract. This paper introduces the *AdRacer* system for multifaceted testing and in-depth analyses of game effects and in-game advertising efficiency. *AdRacer* combines an immersive driving simulator, 3D game environment, recording of players' gaze directions, and application of memory tests. A pilot study tested the effects of game violence on memory for brands shown as billboard ads in a racing game. In contrast to findings with TV violence, game violence did not impede brand memory. Memory results were also not mediated by visual attention during encoding. Compared to a matching nonviolent version, playing a violent game resulted in superior brand retrieval, yet participants showed fewer and shorter eye fixations on the billboard ads. Hence, caution seems to be recommended in transferring standard results from the "passive" TV medium to the interactive game medium.

Keywords: Game violence, in-game advertising, memory, media effects.

1 Introduction

Violent and nonviolent computer and video games have been analyzed with respect to their emotional, behavioral, and cognitive effects (e.g., Anderson & Bushman, 2001). To our knowledge, however, the effects of video game violence on memory for advertised brands have not yet been tested. The present paper introduces *AdRacer*, a novel testing system for the study of game effects. *AdRacer* warrants high levels of experimental control in a realistic and immersive 3D game environment. The system was tested in a pilot study on the effects of playing a violent or nonviolent racing game on memory for brands. Brand information was provided as billboard ads, thus matching what is found on real race tracks. This so-called *in-game advertising* denotes the contextual placement of brands within games and is currently becoming increasingly important for advertisers (section 2). Following a detailed description of the *AdRacer* system (section 3), we will discuss the importance of analyzing effects of game violence on brand memory and report results from the pilot study (section 4).

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 11–22, 2008.

2 In-Game Advertising

In-game advertising has become a major topic because of people's dramatic changes in (entertainment) media behavior. By middle childhood, for example, playing computer and video games has become one of the favorite leisure-time activities (cf. von Salisch, Oppl, & Kristen, 2006), thus rivaling TV as the former anchor medium for entertainment. In addition, media diversification describes peoples' tendency to using various other sources of media entertainment such as mobile phones, PDAs, and, video games. The ever-accelerating economic importance of in-game advertising is also reflected by the numbers. In 2006, companies spent $77.7 million for in-game advertising. In 2007, the budget more than doubled to $182.7 million, whereas in 2011 it is projected that $971.3 million will be spent on in-game advertising (Yankee Group[1]).

Advertisers' step into games is promising because video games grant foolproof access to the coveted target group of young adults who are known to be avid gamers. Young adults also have more expendable income than other groups, and have more malleable attitudes. Second, heavy gamers accept in-game advertising as an inevitable part of the future of their play[2]. Gamers even appreciate "cool" companies that advertise in games[3]. Third, in-game advertising is non-obtrusive. Unlike TV commercials that disrupt a running TV program, advertised brands are carefully and subtly integrated into the game. Most importantly, however, games differ from TV programs and movies in terms of their inherent interactivity. Playing games is an active process, which may even include some form of *forced exposure* to the advertised brand. TV commercials typically entail shallow encoding of product information. Carefully integrating brands as an important part of the storyline may result in deliberately directing the player's avatar to a billboard or even "force" the player to use a particular product (e.g., Coca Cola). From a human memory perspective, this process induces deep encoding of the product information, which, according to the levels-of-processing approach (Craik & Lockhart, 1972), will serve as a superior basis for later retrieval. Taken together, in-game advertising may be among the most effective forms of advertising because of superior brand exposure in terms of frequency, duration, and intensity or depth of encoding.

3 *AdRacer* Testing System

The *AdRacer* system was designed as a flexible tool for adaptive and ecologically valid experimental testing that warrants continuous interaction between the player and virtual 3D game environment. The system comprises driving simulator, 3D racing game environment, and recording of gaze directions and bio signals (e.g., EEG). Additional questionnaires and post-driving memory tests complete the in-depth analysis of advertising efficiency and game effects on the player (Figure 1).

The *AdRacer* is equipped with a 30-inch back-projected display for immersive gaming, an adjustable car seat, stereo sound, and a steering wheel and foot pedals for realistic driving experiences. The system uses a non-obtrusive single-camera infrared eye-tracking device. Figure 2 illustrates the *AdRacer* hardware setup.

[1] http://www.yankeegroup.com/ResearchDocument.do?id=16395, acc. 7/16/2008.
[2] http://www.comscore.com/press/release.asp?press=861, acc. 7/16/2008.
[3] http://www.massiveincorporated.com/casestudies.html, acc. 7/16/2008.

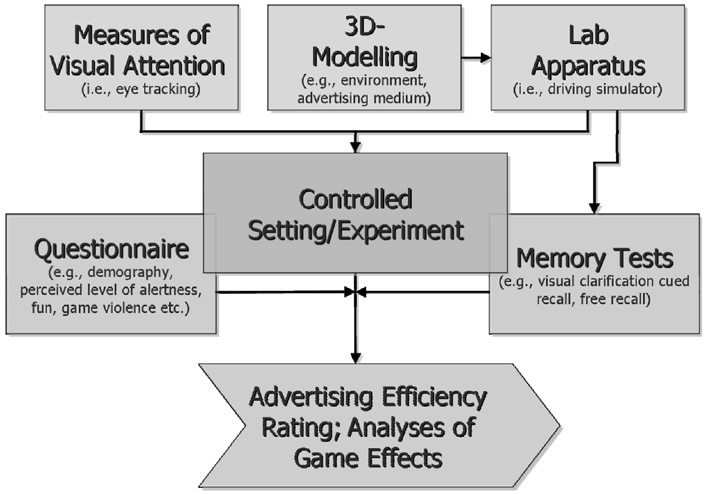


Fig. 1. The *AdRacer* testing system. See text for further details.

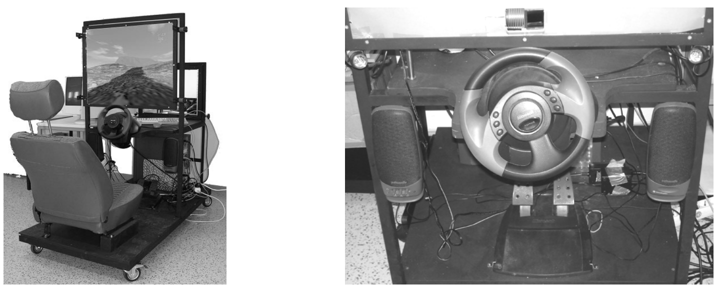


Fig. 2. The *AdRacer* driving simulator, together with steering wheel, stereo speaker, foot pedals, eye-tracking camera (*above steering wheel*) and infrared emitter (*above speakers*)

The virtual 3D environment in the novel *AdRacer* game used in the pilot study was designed using the open source TriBase game engine, which supports Microsoft's DirectX 9 graphics standard and comprises a wide range of I/O and A/V libraries (Scherfgen, 2006). We implemented different 3D models for billboards, trees, houses, rotating geometrical target shapes (i.e., rings and diamonds), and animated human-like characters. For the violent version, two different 3D avatar "targets" were designed (a man in a business suit, and a woman in a wheelchair), both with their hands down. To render a more realistic impression, a second model was designed for each avatar with their hands held up. Once the player approached the avatars, the "hands-up" model automatically replaced the "hands-down" model. This simple two-stage animation proved to be an effective method to induce the lively impression of a "real" person spontaneously reacting to a fast approaching car.

Based on different types of ground tiles (i.e., basic straight, curve-left, and curve-right textures that define the course of the road), and the number and positions of bonus items and billboards, the TriBase engine generated an open race track. The race track is visually situated in the suburbs, with different types of houses and trees sparsely distributed. In the present study, no other cars or AI-based non-player characters appeared (except for the avatars in the violent version). The game's underlying physics engine rendered a realistic driving behavior, including screeching tires and crashes whenever participants left the road and bumped into houses or trees.

Both game versions had a first-person point of view that shifted realistically according to the player's input (Figure 3). By providing "unfiltered" sensory cues, feelings of spatial self-presence are strengthened, thus further inducing involvement and immersion in the game environment (e.g., Ijsselsteijn, 2001). The first-person point of view also plays an important role in current models of media effects. Because of adopting the acting character's role, the first-person view supports identification. In the General Aggression Model (Anderson & Bushman, 2002), these attributes help form chronically accessible mental models, which change an individual's personality and, thus, help to explain how violent video games influence aggressive personality.

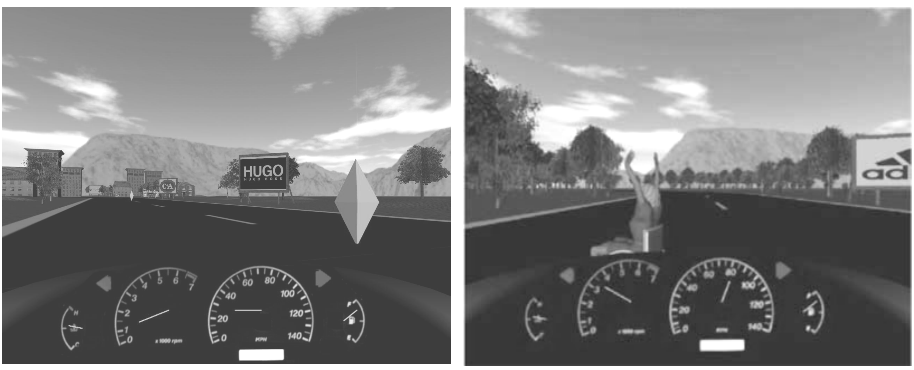


Fig. 3. Hitting items in the nonviolent version (*left*) and running over pedestrians in the violent version (*right*) of the *AdRacer* racing game, together with billboard ads, as seen through the imaginary windshield. An animated dashboard further increased the realistic impression.

To further understand the effects of in-game advertising, an important potential mediator of memory—participants' gaze directions—was also measured. Analyzing potential differences in visual attention towards brand logos helps to understand the mechanisms that underlie brand memory, and gaming behavior per se. Gaze directions provide a continuous measure without interrupting processing. Unlike verbal statements, gaze directions are largely automatic (i.e., resistant to strategic control). Eye movements were continuously recorded in both conditions during the driving game. The *AdRacer* system uses the InSight™ eye-tracking device (©SensoMotoric Instruments), which only requires a headband with a marker tracked by a single camera, thus indicating the player's head position. Based on three infrared emitters, two corneal reflexes are produced giving an estimate of the cornea's curvature. The eye tracker software provides information on head orientation, lid opening, and gaze direction. In the present study, we will confine ourselves to gaze direction. Using a non-obtrusive device further supports the ecological validity of the interaction between player and game environment.

The *AdRacer* display was virtually divided into 42 areas of interest to calibrate individual gaze direction. We programmed a merging routine that combined the recorded signals into one file container, including the AdRacer game data. An additional database handled both file container entries and memory data.

4 Pilot Study: Testing Memory for Product Information

Typically, people watch TV commercials or play sports simulation games in the privacy of their homes, or pass by billboards on their daily way to the office, and then go shopping later. Contact with product information thus occurs prior the actual buying situation. This temporal and spatial decoupling of encoding and retrieval is also reflected in theoretical models of people's complex response to advertising in which memory represents a crucial component (Shimp & Gresham, 1983). Memory has been repeatedly demonstrated to mediate consumer behavior (Bagozzi et al., 1992; Bushman, 2005). Brand placements, for example, are thought to increase the level of familiarity with the advertised brand so that consumers will later remember and eventually buy the product (d'Astous & Chartier, 2000).

In addition to formal aspects (e.g., logo design), memory for advertised brands is also affected by contextual factors. For example, embedding the product information in a violent TV program context is known to impede brand memory compared to a neutral program (see Bushman & Phillips, 2001 for a meta-analytic review). In these studies, participants watch TV with commercial breaks. They are then given surprise tests of memory. Despite comparable levels of arousal, entertainment, and involvement, participants recall fewer brands embedded in the violent TV program (e.g., "24") compared to the nonviolent program (e.g. "America's funniest animals").

In the light of the aforementioned trend towards in-game advertising, the *AdRacer* pilot study tested whether game violence will have detrimental effects on brand memory similar to those found with TV violence. Little is known about brand memory in games; recently, it has even been characterized "a virtually unresearched area" (Yang et al., 2006). The few studies that have already addressed memory for in-game advertising yielded only low brand memory. For example, participants in a first-person shooter recalled going past billboards, yet revealed little memory for brands (Chaney et al., 2004). To our knowledge, no study has directly compared memory for advertised brands in matching versions of a violent and a nonviolent game.

4.1 Method

The experimental pilot study that tested the *AdRacer* system addressed the effects of game context on memory for brands that appeared as billboard ads in a novel racing game—does the violent game context impair participants' memory for brands similar to TV violence? Two matching versions of the game were therefore designed that differed only in terms of violence. In the nonviolent version, participants were rewarded for running over animated geometrical shapes distributed along the race track. In the violent version, participants were rewarded for running over pedestrians (i.e., human-like avatars). Both groups were then given surprise memory tests.

4.1.1 Design and Participants

A 2 (game content: violent, nonviolent) x 2 (memory test: cued recall version of a brand clarification task, free recall of brands) mixed-factorial design was used. Game content served as a between-subjects factor; participants thus either played the violent or the nonviolent version of the racing game. Memory test was varied within subjects. However, the order of the tests was fixed; participants started with the clarification task (cued recall) and then proceeded with free recall.

Participants (N=19) were students and faculty members from different faculties at the University of Luebeck. In the violent condition, 2 females and 7 males (mean age=24.40, SD=4.17) participated[4]. In the nonviolent condition, there were 3 females and 7 males (mean age=22.56 SD=5.41). Faculty members volunteered whereas students received course credit for their participation. Members of both groups were naïve to the experiment and its goals.

4.1.2 Materials

Sixty-four high quality versions of corporate brand logos from different product categories were selected from the Internet[5]. A pretest with 20 different participants corroborated that all logos represented known brands. For the driving game, logo size was adjusted to fit the VR version of the standard German "Mega light" billboard size (252cm x 356cm). For the clarification task that was used as a cued recall test, brand logos were adjusted to fit into 480 (width) by 640 (height) frame size.

All participants were given the same race track in the driving session. However, 32 different logos were randomly selected for each participant. Brand logos were assigned billboard positions that had been pre-rendered by the game engine. Each logo was shown as a 2D-billboard ad and appeared three times during an individual driving session. Hence, each participant encountered a total of 96 brand logos in the study session ("learned list logos"). For the clarification cued recall test, 16 brand logos were taken from the learned list ("repeated") and were presented together with the 16 logos that had not been previously encountered in the driving session ("new")[6]. Repeated and new items were presented in random order. Only one brand logo was displayed at a time in the clarification cued recall test.

4.1.3 Procedure

The experiment comprised, in chronological order, the *training session* to accustom participants to the driving game, the *study session* in which eye movements were recorded while participants played the driving game and encountered billboard ads, the *questionnaire session* that addressed demographical factors and participants' ratings of the *AdRacer* game, and the *memory tests*.

Participants were tested individually in a computer lab. They randomly assigned themselves to either the violent or nonviolent version by drawing a slip of paper. After adjusting the driver's seat in front of the *AdRacer* display the experimenter

---

[4] A third female participant in the violent condition started the driving session but soon had to quit because of experiencing intense simulator sickness.

[5] http://www.webchantier.com/_index_en.html, acc. 7/16/08.

[6] The remaining 16 brand logos from the learned list were used in another memory test that yielded similar results, but will not be reported here.

attached the headband and calibrated the eye tracker. To compensate for individual differences in experiences with driving games, a learning criterion required 180,000 bonus points to be scored within three minutes of driving in the *training session*. A special nonviolent training game version was used. Participants were told to maximize their personal bonus by hitting (i.e., passing through) items that displayed their score and appeared on the left and right lane. Hitting an item triggered a cheerful sound, a reddened screen image (flash) for 200ms, and the bonus score, which was shown in the middle of the screen. No billboard ads were shown during training. All participants met the criterion within three attempts.

Next, both groups were given identical instructions for the driving session (*study phase*); they were told to maximize their personal bonus to make it into the high score list. To further increase motivation, participants were told that the top 3 scorers would receive additional gifts. Billboard ads were not mentioned, nor were participants in the violent condition told that avatars would now replace neutral items. Driving session was self-paced and ended when participants passed by all 96 brand logos.

In the violent condition, "running over" an avatar immediately triggered screams of pain either from a female or male voice depending on the gender of the character. Also, a splashy sound was played and the windshield was covered with blood stains for 200ms. As was true for the nonviolent version, participants were immediately rewarded with the bonus score.

Following the driving session, participants received the *questionnaire*. Next, they started the first *memory test*. They were told that some of the hidden brand logos in the upcoming visual clarification test had been presented earlier on billboard ads. Therefore, they should deliberately recall the driving session. Brands were presented one at a time. Each trial started with a verbal cue displayed for 1,500ms, immediately followed by a fixation cross in the middle of the computer monitor. After 500ms, the fixation cross disappeared and the screen was blackened for 1,000ms. Then, the brand clarification started automatically. The computer program started time recording as soon as a visually degraded brand appeared on the screen. Visual masking was automatically evenly reduced (i.e., gradually clarified) by randomly removing blurring pixels at a 2 percent per second rate, starting with 100% noise blur. Participants were instructed to immediately stop the clarification process (and, thus, recording of reaction time) by pressing the Space bar when they visually identified the brand. The program immediately replaced the degraded logo with the instruction to type in the name of the recognized brand. Participants accustomed themselves to the task on five training trials (using new brands) and then completed 32 test trials.

In the concluding free recall test, participants were given a blank form. They were told to type in only brand names they remembered from the driving session. Finally, participants were debriefed. The entire experiment took 50 to 60 minutes.

4.2 Results

Findings from the questionnaire will be described first (4.2.1), followed by memory data (4.2.2). We will also address whether visual attention (gaze directions) affected brand memory (4.2.3). An exhaustive description of all results would go beyond the scope of this paper. Hence, we will confine ourselves to the most important findings.

4.2.1 Questionnaire

The 20 items of the questionnaire addressed demographical factors, like age, gender, occupation, computer and gaming expertise, and personal frequency of gaming, which are known to moderate measures of gaming behavior. Lower ratings on the 4-point scale indicated stronger affirmation. Two-sided t-tests (α=.05) revealed only one significant effect: game versions substantially differed in violence, $t(17)=6.08$, $p<.01$. As expected, participants in the nonviolent condition (M=3.67, SD=0.50) did not find the game violent at all, whereas players of the violent version (M=1.80, SD=0.79) confirmed that they had played a violent game. No other comparison was significant (ps≥.11), indicating that game versions were matching in terms of handling, perceived realism, and subjective estimations of arousal, excitement, alertness, and fun.

4.2.2 Memory Data

Cued Recall (Clarification Task)

The number of erroneously recalled brand names was below 3%. Therefore errors were not calculated separately. First, we compared baseline performance, that is, group means for brand logos that had *not* been encountered in the study phase. The one-factorial ANOVA with game content serving as a between-subjects variable revealed no group difference, $F(1,17)<1$. Not surprisingly, participants playing the violent version (M=13,276ms, SD=3,006), and the nonviolent version (M=13,548ms, SD=3,105) did not differ in terms of the time it took them to identify blurred brand logos that had *not* been shown earlier. Next, we compared group means for brand logos that were *repeated* from the study phase. Contrary to our hypothesis, the one-factorial ANOVA revealed no group difference, $F(1,17)<1$. Participants playing the violent version (M=12,878ms, SD=2,894) were even numerically faster than their colleagues in the nonviolent version (M=13,614ms, SD=2,923).

Free Recall

There was no substantial group difference in the free recall test, $t(17)=-.095$, $p=.93$. In the nonviolent condition, participants recalled 9.03% (M=2.89, SD=2.09) of the brand logos encountered in the driving session. Participants playing the violent version recalled 9.38% (M=3.00 SD=2.91).

4.2.3 Eye-Tracking Data

Due to technical problems with the eye-tracking device, data from only 5 participants in the nonviolent and 8 participants in the violent condition were obtained. Because of the low number of observations, only descriptive statistics will be presented. First, the mean number of fixations (*hits*) for billboard ads was calculated (Table 1, upper half). Recordings from gaze directions up to 3% displacement from the outer billboard frame were accepted as hits[7]. Please note that each of the 32 different brand logos was repeated three times during an individual driving session.

Apparently, repeating brand logos as billboard ads in the driving session positively affected participants' gaze directions only in the nonviolent game version. In the violent version, visual contacts with billboards occurred only at chance level, irrespective of their number of appearance.

[7] We also tested different displacement criteria (0%, 1%, and 5%), but obtained similar results.

Table 1. Mean number of eye fixations (in %, *upper half*) and mean duration of eye fixations (in ms, *lower half*) in the two game versions. Standard deviations are given in parentheses.

		Number of Appearances		
		First	**Second**	**Third**
Number of Hits	**Nonviolent**	42.50 (36.28)	56.25 (26.88)	68.75 (12.50)
	Violent	56.25 (21.65)	51.56 (26.67)	52.34 (35.82)
Fixation Times	**Nonviolent**	286 (268)	234 (323)	276 (320)
	Violent	166 (248)	192 (265)	167 (263)

Next, mean fixation time for "hits" was calculated (Table 1, lower half). Playing the violent game version not only resulted in numerically fewer, but also shorter eye contact with billboard ads. Compared to the nonviolent version, thus, "violent" gamers generally seem to pay less visual attention to ads. In addition, the expected correlation between fixation times and number of hits was significant in this condition, r=.82, p=.01, but not in the nonviolent version, r=-.22, p=.73.

In a final step, eye-tracking data and memory results were compared. To our surprise, there was no significant correlation between gaze direction and memory performance for repeated brands in the later cued recall (clarification task). In the nonviolent game version, additional (r=-.24) and longer eye contact (r=-.35) with billboard ads only numerically speeded later identification of repeated brands. In the violent condition, however, the positive yet insignificant correlation of number of hits (r=.22), and fixation time (r=.39), respectively, indicates that eye contact with the brands even numerically slowed down brand identification.

4.3 Discussion

The *AdRacer* pilot study yielded surprising findings. Results from both memory tests sharply contrast with our hypothesis based on previous findings that demonstrated the detrimental results of TV violence on memory (e.g., Bushman & Phillips, 2001)[8]. In our study, game violence (i.e., running over pedestrians in a racing game) did not at all impede memory for brands previously shown as billboards. We found no significant differences on either memory test between violent game players and nonviolent game players. Visual attention did not mediate the results either. How do we make sense of these findings?

Most importantly, only a small number of participants were tested in the *AdRacer* pilot study. Technical problems with the eye-tracking device further reduced the number of observations. First and foremost, we thus have to replicate our findings with a larger sample to increase the reliability of our findings.

---

[8] In the literature on media effects on brand memory, two studies have reported a similar result (i.e., greater memory performance in the violent condition). The authors of the first study did not provide a cogent explanation for their findings (Droulers & Roullet, 2004). The second study showed that reinstating a "violent mood" at test might support retrieval of violent commercials (Gunter et al., 2005). In our study, however, brands were emotionally neutral and memory was tested with an emotionally neutral visual clarification task and free recall.

In addition, one could argue that game versions did not differ in level of violence they conveyed. As expected, however, participants' ratings confirmed that collecting geometrical shapes and running over pedestrians significantly differed in terms of violence. Although subjective ratings do not fully rule out the possibility that versions might have been too similar—we did not record objective physiological data to indicate typical patterns of bodily reaction to violence, for example—we are confident that we have successfully varied game violence in the present study.

The detrimental effect of TV violence on brand memory has been attributed to violence drawing visual attention away from the product information to the program itself, thus preventing brand memory traces from being established (i.e., attention-grabbing effect; Bushman & Bonacci, 2002; Gunter et al., 2005). This thesis is also supported by recent models on the effects of emotional arousal on memory binding (Mather, 2007). Attention is assumed to be primarily paid to the emotionally arousing object. At the same time, contextual features are neglected and therefore not properly encoded. In our study, we expected participants playing the violent racing game to pay attention to the task, and, thus, to their potential victims (i.e., the emotionally arousing objects), but not to task irrelevant contextual features like billboards. Interestingly, "violent players" in fact dominated the high score list in the driving session; six of the top ten players (including ranks 1 and 2) played the violent version with a mean score of 1,212.500 (mean score for "nonviolent players" was 1,199.875). Likewise, eye-tracking data indicated that participants in the violent version did not benefit from repeating brand logos. Apparently, participants closely *did* pay attention to the driving task.

But if playing the violent version indeed meant paying close attention to violent acts, but not to brand logos, why did participants in the nonviolent condition perform so poorly in the memory tests? One attempt we made to match both conditions as closely as possible could have backfired on the study, selectively affecting the nonviolent version. To compensate for the blood-covered windshield in the violent condition, passing through a geometrical shape triggered a flash of red light for 200ms. In contrast to the semi-transparent blood splashes, the red flash was opaque and entirely covered the screen. Hence, a masking effect may have disrupted visual perception in the nonviolent condition. Visual masking is an effective way to erase iconic memory, thus preventing consolidation of representations (Enns & Di Lollo, 2000). After the mask disappeared, participants in the nonviolent condition may have "re-fixated" by means of additional eye movements. This post-hoc explanation would fit both memory and eye-tracking data (i.e., additional fixations and greater fixation time) in this condition and could easily be tested in the next study.

5 Concluding Remarks

The present paper introduced the *AdRacer* system for systematic analyses of game effects. *AdRacer* offers high levels of experimental control in a realistic and immersive 3D game environment. Because of the system's underlying architecture, gaming information may also be utilized in subsequent tests of cognitive processes (e.g., memory). By integrating additional methods of detecting the player's gaze direction and physiological data (e.g., EEG), *AdRacer* also supports the in-depth analysis of actual gaming behavior. Further analyzing the processes engaged during game play is an indispensable prerequisite of understanding the effects or consequences of game violence on players' reactions to brands.

AdRacer was used in a pilot study that tested the effects of game violence on brand memory. Remembering product information is becoming increasingly important in the upcoming economic efforts of in-game advertising, yet little is known about its effects (Yang et al., 2006). The overall low number of participants certainly limits the reliability of our findings. However, recordings of gaze directions and analyses of brand memory revealed two remarkable results. First, encountering task irrelevant brand information may establish memory traces that are later reenacted. This is an important finding for advertisers thinking of product placement in computer and video games. However, advertisers should also note that this memory effect was far from being irreversible. Rather, memory was sensitive to contextual factors. At this point, we cannot specify whether this sensitivity was due to game version (violent or non-violent), or short visual flashes. Future studies will further our understanding if game violence will mirror the detrimental effect of TV violence on brand memory, which has been repeatedly demonstrated (e.g., Bushman & Phillips, 2001). Given the increasing economic relevance, the important role of memory processes in the context of advertising, and the societal implications of violent games, future studies are badly needed. The AdRacer provides an excellent platform for conducting this research.

Acknowledgments

The authors wish to thank Olaf Christ, Sebastian Otto, Henning Hansen, and Gregor Radzimski for their pivotal programming work and running the experiments.

References

1. Anderson, C.A., Bushman, B.J.: Human aggression. Annual Review of Psychology 53, 27–51 (2002)
2. Anderson, C.A., Bushman, B.J.: Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior: a meta-analytic review of the scientific literature. Psychological Science 12(5), 353–359 (2001)
3. Bagozzi, R.P., Baumgartner, H., Yi, Y.: State versus action orientation and the theory of reasoned action: An application to coupon usage. Journal of Consumer Research 18, 505–518 (1992)
4. Bushman, B.J.: Violence and sex in television programs do not sell products in advertisements. Psychological Science 16(9), 702–708 (2005)
5. Bushman, B.J., Bonacci, A.M.: Violence and sex impair memory for television ads. Journal of Applied Psychology 87(3), 557–564 (2002)
6. Bushman, B.J., Phillips, C.M.: If the television program bleeds, memory for the advertisement recedes. Current Directions in Psychological Science 10, 44–47 (2001)
7. Chaney, I.M., Ku-Ho, L., Chaney, J.: The effect of billboards within the gaming environment. Journal of Interactive Media 5(1), 54–69 (2004)

8. Craik, F.I.M., Lockhart, R.S.: Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior 11, 671–684 (1972)
9. d'Astous, A., Chartier, F.: A study of factors affecting consumer evaluations and memory of product placements in movies. Journal of Current Issues and Research in Advertising 22(2), 31–40 (2000)
10. Droulers, O., Roullet, B.: Does crime pay for violent program-embedded ads? Advances in Consumer Research 31, 1–6 (2004)
11. Enns, J.T., Di Lollo, V.: What's new in visual masking? Trends in Cognitive Science 4, 345–352 (2000)
12. Ijsselsteijn, W.A., Freeman, J., De Ridder, H.: Presence: Where are we? CyperPsychology & Behavior 4(2), 179–182 (2001)
13. Gunter, B., Furnham, A., Pappa, E.: Effects of television violence on memory for violent and nonviolent advertising. Journal of Applied Social Psychology 35(8), 1680–1697 (2005)
14. Mather, M.: Emotional arousal and memory binding. An object-based framework. Perspectives on Psychological Science 2(1), 33–52 (2007)
15. von Salisch, M., Oppl, C., Kristen, A.: What attracts children? In: Vorderer, P., Bryant, J. (eds.) Playing Video Games: Motives, Responses, and Consequences, pp. 147–163. Erlbaum, Mahwah (2006)
16. Scherfgen, D.: 3D-Spieleprogrammierung mit DirectX 9 und C++. Hanser, München (2006)
17. Shimp, T.A., Gresham, L.G.: An information processing perspective of recent advertising. Current Issues and Research in Advertising 9, 36–79 (1983)
18. Yang, M., Roskos-Ewoldsen, D.R., Dinu, L., Arpan, L.M.: The effectiveness of "in-game" advertising. Comparing college students' explicit and implicit memory for brand names. Journal of Advertising 35(4), 143–152 (2006)

Immersion, the Greatest Hook

Faisal Qureshi

Devry University
fqureshi@devry.edu

Abstract. People play certain genres of video games for definitive reasons. The underlying phenomenon of these manifested reasons, and the source of ”hooking” a player to a game, is Game/World Immersion. Immersion, therefore, is the single most important factor as to why people play certain genres of video games. The various elements that are present in such games, such as audio or visual, are there to create the sense of immersion and are the building blocks of it.

1 Introduction

There is a driving force that compels certain people to play 3rd person action/adventure, role-playing, and first person shooter games. These gamers, hardcore or casual, play for definitive reasons. The underlying reasons why people are drawn to interactive digital entertainment are escapism, stress-relief, challenge, and socialization. These motivational factors are linked to a single phenomenon, for these genres, that manifests itself as these categories. This phenomenon is Game/World Immersion.

2 Design Factors

The sense of game immersion is created by multiple elements. The audio and visuals of the game are the obvious ones. The more realistic the game looks, the greater the immersion. The reason why we have evolved to 3-dimensions is to enhance the realism, and ergo the immersion. With the addition of balanced game-play challenges, the player is further drawn into the virtual world. The actions that are required to overcome these challenges give life to the virtual world and once again, enhance the immersive experience.

(1) Many successful immersive video games lead the player toward completion of pre-planned goals, just like novels or movies. Ideally, when those goals are accomplished and the player is rewarded for those specific successes, the player should feel as though it could not have been accomplished without his skill Steve Woyach (Dunniway 66,81).

(1) Unfortunately, this method presents a dichotomy for the player, because the elements of story and gameplay are naturally separate. Thus when the two are intertwined in a game, the player often feels that either the story is an unwanted distraction from the gameplay or that the gameplay is a tedious drawing-out of the story. However, the evolution of video games has resulted in a new technique for integrating story and gameplay more intricately than in the past. Steve Woyach.

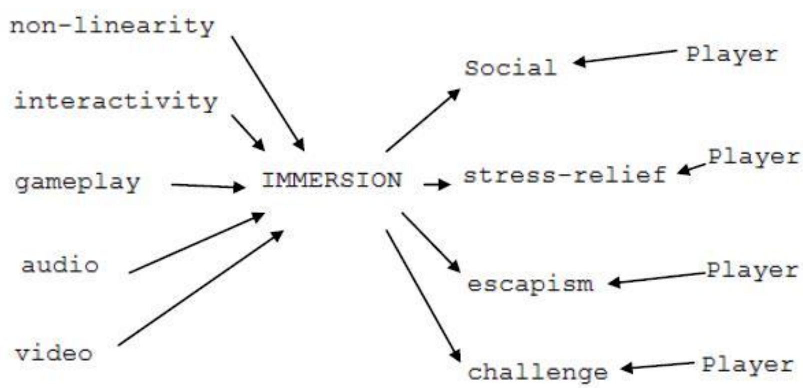
S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 23–26, 2008.

This implies that due to the phenomenon of interactivity in games, players are dropped directly in the virtual and can perform actions in it, either interacting with the world, or on the world. Furthermore, gameplay elements, and as mentioned by Steve Woyach, are directly connected with the story, both which create a cohesive environment and an enhanced immersive world. So not only is the look of the game a building block for immersion, but the feel and gameplay also are factors that add to an overall immersive experience.

The non-linearity that interactivity creates, as well as the audio and visual of a game, creates the look and feel of the game which in turn immerses the gamer into the virtual world. The story of the game is also a critical element of world immersion. Life is a story about oneself from the individuals perspective. Through video games, we can virtually live out other possible paths by choice and interaction. We choose which path we want to live via the game the user decides to play.

(1) This results in a tremendously complicated design and programming process, but can give rise to an unparalleled range of freedom in terms of gameplay. The player can choose from in infinite number of strategies, because their imagination is the only limit to how they can proceed. As in life, we are given a set of abilities, skills, and procedures, and we are free to use them as we will. Thus, a giant leap is made towards immersion.-Steve Woyach

The following diagram illustrates the building blocks of immersion, the different manifestations of immersion, and the connection with the player:



3 The Avatar

Furthermore, a high ranking game element that creates immersion is the avatar. You play as the avatar in a game and consciously or unconsciously assign your identity to that main character. In doing so, you have actually placed yourself directly in the virtual world. This would not be possible without some sort of avatar. Since it is very rare in 3rd person action/adventure, FPS, or RPGs, to exist without an avatar, the game itself, by default, is an immersive experience by virtue of having an avatar. Some games allow you to customize the avatar which causes an extra level of the player attaching his/her identity and personality to that avatar.

(2) This connection between the player and the players position in the game space implies a type of identification, in that the player identifies sufficiently with objects or characters of the game space to function in response to that game space through a self-image that is inserted into the constructs of the game space and then internalized by the player (subject) Laurie Taylor.

In her article, When Seams Fall Apart Video Game Space and the Player, Laurie Taylor discusses a concept called Telepresence where the player exists in multiple places at the same time. This means the player not only exists in physical space, but in the virtual world as well. This is simply another way of saying that the player is immersed in the virtual world as well as exiting in physical reality. Taking the example of the God of War series, the avatar is a Spartan named Kratos. The connection to Kratos is a psychological will for great strength and power. In the game, the player can feel the strength of the character through various actions and challenges. By feeling what the character experiences, the player has connected with the avatar and since the avatar is in a virtual world, the connection also places the player in the virtual world. For other avatars, players make different connections with the avatar for different psychological or emotional reasons. In the end, you can be anything you fantasize about as long as the avatar permits it.

4 Manifestations

If escapism is one reason why people play games, then where are they escaping to? The obvious answer is the virtual world. This concept is synonymous to compelling immersion. If people play games to relieve stress, the question becomes how. The game provides an outlet from mundane issues, and by the very nature of the word outlet, we are implying virtual immersion. If someone is playing a game simply to socialize via technology, a philosophical concept arises of whether this is still a virtualization of reality. For example, when you talk on the phone, you hear to voice of the person on the other end, but how do you know whether he or she is really there? This concept is similar to Schrodingers cat. If the conclusion is that this scenario exists only virtually, then immersion, yet again, has triumphed.

5 Conclusion

From what has been discussed, creating an interactive immersive experience seems to be the primary goal of 3rd person action/adventures, first-person shooters, and role-playing games, and the underlying reason why people play these genres. If this is indeed true, then by logical deduction becomes the first and greatest hook of this form of media for these genres. Hooks that exist outside of immersion are then simply different manifestations of the source, which is immersion. The argument that the multiple factors that compose the game itself are directly connected to immersion, would make the existence of these genres without immersion a contradiction, and would negate such existence. This leads me to believe that since people do in fact play these genres, and these games root is immersion, people buy and play games, hence being hooked, foremost to be dissolved in a virtual world where the spectrum of possibilities are only limited by the games creators imagination. As games evolve, an interactive motion capture game with a 360 degree display will not reveal itself unless designers are attempting to create an ultimate immersive experience.

References

1. Woyach, S.: Immersion Through Video Games. Illumin 9(ii)
2. Taylor, L.: When Seams Fall Apart Video Game Space and the Player. Game Studies, the international journal of computer game research 3(2)

Newsgames: Theory and Design

Miguel Sicart

Assistant Professor
Center for Computer Games Research
IT University of Copenhagen
miguel@itu.dk

Abstract. Computer games have a long history as entertainment media, but their use for educational or political communication is relatively recent. This paper explores the use of computer games as news media. Newsgames are computer games used to participate in the public sphere with the intention of explaining or commenting on current news. The paper provides a set of concepts for analyzing newsgames, based on public service theory. The paper expands this analytical approach with a reflection on game design methodologies for creating newsgames.

Keywords: Newsgames, Serious Games, Public Service Theory, Procedural Rhetorics, Political Games.

1 Introduction

In November 2007, the Danish electoral campaign was at its peak. There were political messages in all mass media, from the press and television to Internet sites and social networks. But the 2007 Danish elections will be remembered as the first elections were most political parties developed and launched a computer game.[1]

The path from *SpaceWar!* to political games illustrates the cultural impact of information technologies and the Internet in Western Societies. Fast, networked computers have become as ubiquitous as television or radios, and software for game development has become easier to learn and distribute. It is a natural step to use these favorable conditions for developing games that answer to deeper needs than entertainment. Computer games can also participate in public debates, by means of their procedural rhetorics [1].

I will define newsgames as those games that utilize the medium with the intention of participating in the public debate. The ease of development and distribution contribute to the evolution of game that comment or simulate news events, understood as "the preexisting discourse of an impersonal social institution which is also an industry"[2].

[1] The Danish journalist Thomas Vigild commented these games on his blog: http://blog.politiken.dk/klik/category/politik/. Political games are not new: the 2004 United States election saw the first candidate-endorsed game, *Dean for Iowa* (http://www.deanforamericagame.com/), and in 2005 Newsgaming.com developed *Cambiemos*, a game for the Uruguayan socialist party (http://www.ludology.org/my_games.html).

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 27–33, 2008.

The term "newsgame" lacks a formal definition applicable both for analysis and design. I will define newsgames using public service theory and concepts from computer game studies, thus providing a general framework for the analysis of newsgames.

A second goal is to present newsgames in relation to computer game design. I will reflect on methods for designing newsgames, applying the theories on value-based design by Nisembaum, Flanagan, Belman and Diamond [3] and the work on computer game rhetorics by Bogost and Frasca [4].

In this paper I will argue for computer games as a valid tool for public debates and exchanges of ideas between citizens: as a civil tool for simulation and communication.

2 Newsgames: Theory

Serious games are games used for different purposes than just entertainment [5]. More precisely, serious games use the rhetorics of games [6] to convey messages with intentions different than to entertain. This category potentially includes all kinds of games, from classic board and role playing games to simulations on computers.

In this paper I focus on serious computer games, as defined by Bergeron [7]: "an interactive computer aplication, with or without a significant hardware component, that: has a challenging goal; is fun to play and/or engaging; incorporates some concept of scoring; imparts to the user a skill, knowledge, or attitude that can be applied in the real world" (xvii). This meta-category encompasses all kinds of serious games. Defining newsgames requires a more focused approach.

Newsgames are serious computer games designed to illustrate a specific and concrete aspect of news by means of their procedural rhetoric, with the goal of participating in the public debate. Newsgames are ephemeral as the news they illustrate, and they often have editorial lines correspondent to the lines dictated by their parent media.

In this sense, newsgames do not cover general political topics, like *September 12th* [8], nor they are used for political campaigning, like *Dean for America*. Newsgames are designed to accompany a specific piece of news. This implies that newsgames do not have the intention of surviving in the collective memory for a long time, nor they have a replayability requirement. Newsgames illustrate and comment the news, and are as perishable as the news themselves.

Nevertheless, newsgames have the ambition of transmitting important notions about news events by means of their rhetoric. In this sense, public service broadcasting theory casts some light on how media are used to participate in the public debate. I am not implying that newsgames are public service media. Public service broadcasting theory offers a number of arguments that explain the use of traditional media for conveying news; this theory can be expanded to cover newsgames, both produced by private firms and by public corporations.

Public broadcasting is relevant because "broadcasting should have objectives other than the entertainment of viewers and listeners and the profitability of private broadcasting firms" [9]. The same applies to newsgames: they are not for direct profit, and they want to go beyond the limitations of pure entertainment, challenging their audiences to different experiences.

Newsgames are a sub-type of political games because they do not necessarily convey a political message, but the implications of a piece of news. Any transmission of news is editorialized, but newsgames do not need to be received exclusively as political opinion pieces: "people do not normally believe that news exists primarily to influence them. They also do not believe that they consume news for the purpose of being influenced" [10]. Newsgames are a type of political games that do not have as the main goal to transmit a political idea, but to simulate certain aspects of news as relevant for an intended audience. That simulation can be political, but does not need to.

This audience is not understood as consumers, but as citizens[2] that share a vision of the public interest. Denhard & Denhard [11] define the role of public administrators as that of "building a collective, share notion of the public interest", and it is in this perspective newgames are tools for shaping the public space for debate on a specific topic.

Having citizens as target group implies that newsgames do not intend to convince people about a specific agenda related with news, but to convey a space of opinion within a space of possibility, defined as "the space of future action implied by a game design" [12]. Unlike the broader category of political games, which try to change their user's frames and ideas on specific political topics,[3] newsgames "serve rather than steer", helping "citizens articulate and meet their shared interests, rather than to attempt to control or steer society in new directions" (Denhard & Denhard, p. 553).

The number of non-steering newsgames is rather small: titles like *Madrid* [13] convey an ideological message, and present other perspectives on news events to the public opinion. These newsgames present a conceptual interpretation of an event that can intervene, and potentially enhance the public debate, without steering the argumentation. *Madrid*, for example, illustrates the collective sense of empathy the bombings in Spain created, but also reflects on the need to actively not forget about this bombings, or any bombings, so history will not repeat. These ideas are presented both in the audiovisual level, as graphics and sounds, and in a systemic way, as rules and processes the player has to interact with in order to experience the game.

This definition of newsgames provides them with a goal. Unlike political games, that present a clearly biased argumentation with the intention of shifting or reassuring the audience in their political beliefs, newsgames do not enforce directly instrumental goals. Newsgames will not change ideas, but provide arguments for engaging in discussions within the public sphere. The goal of newsgames is to address citizens more than consumers: "in that lies the real clue to the nature and purpose of great broadcasting: that it makes best sense when it represent a national and moral optimism within a society, when it suggests (…) that we can be better than we are: better served, better amused, better informed, and thus, better citizens".[14]

Newsgames have a political agenda and an editorial line, but those are presented as an open space for discussion. Political and persuasive games appeal to the partisan, newsgames appeal to the citizen. But newsgames are not, and should not be ideologically neutral. Like any other form of discourse addressed to the public, these games have an agenda, which is presented not as truth, but as arguments. Newsgames are computer games that engage in a public discussion, articulating discourses using procedural rhetoric.

---

[2] "Serve citizens, not consumers", Denhard & Denhard (2000), p. 565.

[3] Frames as defined by Bogost (2007), pp. 99-121.

In 1985 the British Broadcasting Research Unit suggested eight principles that should be present in all types of broadcasted public services[4]. Newsgames are a different media in a different era, yet it is interesting to formulate some principles that ought to be present in newsgames. And so, newsgames (tend to) follow these principles:

- Easy, almost universal distribution: browser based games are often the vehicle for newsgaming, while downloads are more rare and less desired, since they require specific hardware configurations.
- Newsgames can have an editorial line, but they do not have political interests. Newsgames illustrate, not persuade.
- Newsgames participate in the public debate illustrating news by means of procedural rhetorics, but they do not want to steer the discussion.
- Newsgames should only refer to specific news in a specific period of time: newsgames are temporal, they do not aspire to survive longer than the news.

Newsgames are produced with the intention of participating in the public sphere with arguments that illustrate perspectives on news. Those arguments are made by means of the techniques of games: the simulation of systems by means of procedural rhetorics. Newsgames adapt the content of news to the expressive capacities of computer games, creating ephemeral games.

3 Newsgames: Design

Computer games are complex systems for interaction that require careful design to be engaging, balanced and operative. Game design is the craft of creating interesting gameplay with the goal of engaging players in a ludic experience. In computer game productions, most of the design is often done during preproduction, and it can take up to weeks or months to design a computer game.

Newsgames do not have that privilege: they have to be produced and launched while the news are still relevant, not only to participate in the public debate, but also for the game to have any meaning. Newsgames have very focused designs that tend to afford narrow but deep interaction, playing with game design conventions and genres that allow for faster implementations.

But before detailing the design constraints and paradigms for creating newsgames, it is necessary to mention that newsgames, like any other computer game, can have embedded values in their design, as Flanagan et al. have argued [3]. Any game presents a set of values hardcoded in the system design, and presented to players as affordances and constraints [15]. Most of these do not have explicit ethical, political or social goals: commercial games do not put in the center of its experience the expression of interesting dilemmas by means of the interaction with their game system, even though they can be experienced and analyzed as moral or political objects.

Political games and newsgames, on the other hand, make these affordances and constraints an explicit part of their design. Newsgames use these affordances to

[4] These principles are: geographic universality; funding by the (viewer/listening) audience,; independence from government or vested interests; concern for national identity and community; catering for all interests and tastes; catering for minorities; quality of programming; and creative freedom for program makers (Brown: 1996, p. 4; see also Tracey: 1998, pp. 26-32).

convey they meaning of the news, to illustrate their positions and to communicate to the player.

This is, the first design criteria for newsgames: the message should not be hidden. The game as a rhetorical device must be aware of its interpretation of the news, which should also be made explicit and clear to the player. Political and propaganda games can hide the fact that they limit player interaction with expressive purposes; newsgames, in order to be editorial but not steering, need to clearly communicate these design conditions to the player.

The main challenge that designing newsgames poses is translating the content and relevance of news into interesting gameplay that can be experienced by players. This game system has to steer and illuminate the news it is addressing. This means that the game adapts the elements of the news to playable elements that are relevant for understanding the simulated.

News, for game design, should be understood as design constraints. Designing newsgames is translating those constraints into game rules, mechanics, and challenges. Game rules are the basic conditions for the game, including the end state or the winning condition, plus the limits of the gameworld and/or the game experience. Game mechanics are the designed actions players can take within that game experience. In a newsgame, the main news sets both the rules, specially in terms of the winning condition, and the game mechanics, or how the player interacts with the gameworld.

In *Madrid*, for example, players only have one mechanic available: to light candles. They invoke it by clicking with the mouse on the candles, and the rules of the game state that if there is no player input after a certain time, the game loop will stop and the endscreen will be presented. The game has no winning condition, it only has rules that determine both the endstate and the losing condition. By giving players an never-ending task, *Madrid* conveys its message: memory is an active effort, and in the wake of a tragedy, we must also act to remember. This message is conveyed by manipulating game mechanics and conventions, such as the absence of a winning condition.

The mode in which news are translated into game systems determines the editorial line of the game, as it conditions which elements are going to be simulated and which are left out. The ideology of a newsgame, then, can be found in the way the original topic of the news has been translated into the game system, and how that system is presented to a player by means of afforded and constrained mechanics. Returning to *Madrid*: the game translates the difficulties of collective mourning and historic memory by forcing players to a perpetual action. Players are encouraged to remember, even though there is no victory; players are also punished for their passitivity. Unlike *September 12th*, which could only be won by not playing, *Madrid* can only *not be lost* by playing it.

Since production cycles are rather brief, newsgames have very basic game mechanics inspired, or directly copied from classic games. Far from being a problem, this is an advantage for newsgame design. For an game, it is fundamental that players understand the principles of play. In more conventional games, the learning curve is steep and requires some time. But newsgames' players need to be presented from the outset with elements they find familiar. Typically, players of newsgames will not use much time learning new controllers or mechanics, so the access point has to be relatively well know for the average target group user. Thus the use of game conventions: appealing to the player repertoire [16] eases the access and consumption of the game by the intended audience.

On the other hand, newsgames appeal to reflective players that can suspend their own immersion in the game, and see its inner workings as a piece of editorial content. In a phenomenological sense, newsgames requires players to experience play and to experience the experience of play: players need to understand from a meta-level what interaction and simulation means for a particular game; they need to be alienated from conventions so that the message of the game is communicated. For example: This player "alienation" can be done my means of transforming the rules and the game mechanics of a game, as *Madrid* does with the absence of a winning condition.

The rules of a game are designed to create an satisfactory winning condition that is challenging to achieve, and every element of those rules is optimized to create a specific ludic experience in the player. In newsgames, on the other hand, rules and mechanics are the systemic embodiment of the boundaries of the editorial line. Rules are a fundamental part of the meaning of a game as a newsgame. This means that, in newsgames, rules are the opinion: the boundaries where play takes place, those that decide what is simulated, and how. Game mechanics communicate the editorial line of a game by funneling player agency towards a specific type of behaviors relevant for understanding the message of the game. The goal of the newsgame designer is to create a set of mechanics that allow interaction with the world, but are sufficiently opaque so players reflect about the meaning of their actions.

In summary, newsgames translate editorial lines into game mechanics and rules. Salen and Zimmerman have pointed out that computer games tend to obscure the player from the inner workings of the game system.[5] This black box technique blinds players from the algorithimical relations between input and output, and is often used to enhance the alleged immersive capacities of the game, making players focus exclusively in their agency and presence within the gameworld.

Newsgames are more keen on showing their inner system. Newsgames require reflective players that can see the semantics of the game mechanics and rules. To attract and help these players, newsgames present either self-referential mechanics, or mechanics designed for breaking the players' expectations. They key element in newsgame design is not only to translate the news into a game, but also to make players realize the editorial line of the game.

4 Conclusion

In this paper I have argued that computer games, thanks to its unique rhetorical capacities, can be effectively used to participate in the public debate. Newsgames are an editorialized, not persuasive genre of political games or serious games, developed in connection to a specific piece of news.

Newsgames are a hint of the future of computer games, not in terms of technological development, but in their role and influence in our culture. Not using computer games for engaging, and even creating public debate to the illustrated digital public is missing an opportunity for making history with a genre. The use of computer games

[5] See Salen and Zimmerman (2004), pp. 88-89.

to illustrate and discuss must be an imperative in the future developments of the society of information.

References

[1] Bogost, I.: Persuasive Games. The Expressive Power of Videogames. The MIT Press, Cambridge (2007)
[2] Hartley, J.: Understanding News. Routledge, London (1982)
[3] Flanagan, M., Nissembaum, H., Belman, J., Diamond, J.: A Method for Discovering Values in Digital Games. In: Situated Play. Proceedings of DIGRA 2007 Conference, pp. 752–759 (2007)
[4] Frasca, G.: Videogames of the Oppressed: critical thinking, education, tolerance and other trivial issues. In: Wardrip-Fruin, N., Harrigan, P. (eds.) First Person, New Media as Story, Performance, and Game. The MIT Press, Cambridge (2004)
[5] Abt, C.: Serious Games. University Press of America, New York (1970)
[6] Sutton-Smith, B.: The Ambiguity of Play. Harvard University Press, Cambridge (1997)
[7] Bergeron, B.: Developing Serious Games. Charles River Media, Hingham (2006)
[8] `http://www.newsgaming.com/games/index12.htm` (September 12)
[9] Brown, A.: Economics, Public Service Broadcasting, and Social Values. The Journal of Media Economics 9(1), 3–15 (1996)
[10] Gunther, A.C., Thorson, E.: Perceived Persuasive Effects of Product Commercials and Public Service Announcements. Communication Research 19(5), 574–596 (1992)
[11] Denhard, R.B., Denhard, J.V.: The New Public Service: Serving Rather Than Steering. Public Administration Review 60(6), 549–559 (2000)
[12] Salen, K., Zimmerman, E.: Rules of Play. The MIT Press, Cambridge (2004)
[13] Madrid, `http://www.newsgaming.com/games/madrid/`
[14] Tracey, M.: The Decline and Fall of Public Service Broadcasting. Oxford University Press, Oxford (1998)
[15] Norman, D.: The Design of Everyday Things. Basic Books, New York (1988)
[16] Juul, J.: Half Real. Videogames between Real Rules and Fictional Worlds. MIT Press, Cambridge (2005)

Agents That Relate: Improving the Social Believability of Non-Player Characters in Role-Playing Games

Nuno Afonso and Rui Prada

IST-Technical University of Lisbon, INESC-ID
Avenida Prof. Cavaco Silva – TagusPark
2744-016 Porto Salvo, Portugal
{nafonso,rui.prada}@gaips.inesc-id.pt

Abstract. As the video games industry grows and video games become more part of our lives, we are eager for better gaming experiences. One field in which games still have much to gain is in Non-Player Character behavior in socially demanding games, like Role-Playing Games. In Role-Playing Games players have to interact constantly with very simple Non-Player Characters, with no social behavior in most of the cases, which contrasts with the rich social experience that was provided in its traditional pen-and-paper format. What we propose in this paper is that if we create a richer social behavior in Non-Player Characters the player's gaming experience can be improved. In order to attain this we propose a model that has at its core social relationships with/between Non-Player Characters. By doing an evaluation with players, we identified that 80% of them preferred such system, affirming that it created a better gaming experience.

Keywords: role-playing games, non-player characters, artificial intelligence, social, behavior, relationship, personality, theory of mind.

1 Introduction

As computers became more present in our lives, they transformed from simple work machines into a means of entertainment. This created a massive industry of video games, an industry that beats sales records every year, and has a yearly growth rate higher than any other entertainment industry [2].

With the appearance of this industry, many of the games were converted from their traditional form into a digital form. One of these cases was the Role-Playing Game (RPG). In its pen-and-paper version, the RPG has a very big social component, because the players create the story as they interact with each other. However, in its transition to the digital format, the RPG was transformed, in most cases, from a multiplayer into a single-player experience where instead of interacting with other people, the player interacts with autonomous Non-Player Characters (NPCs). In computer RPGs players take the role of portraying a character and goes through the game world resolving tasks that are given to them by NPCs. The player needs to interact with

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 34–45, 2008.

NPCs in order to improve her/his character and progress in the story, which means that even in the digital format, the RPG still has a very big social component.

The problem is that the social component of computer RPGs is yet far from the richness of its pen-and-paper format. Usually, the NPCs are not socially deep and most of the times are simply frozen in time, repeating the same action, if any, over and over again. This fact breaks the social experience of user that in turn may break his suspension of disbelief, which is one of the major factors to achieve a successful gaming experience.

In this paper we propose a simple model that can improve the social deepness of NPCs. It is centered on the idea that NPCs establish relationships with others (including the player). Moreover, NPCs have unique personalities that together with the relationships they establish and their knowledge of the relations of the others (e.g. through a theory of mind) influence their behavior. The behavior is reflected in the dialogs NPCs generate, which among other things include the quantity of information they give to the user. The game dialogues are, therefore, adapted to the social context of the game.

The paper is organized as follows. First, we describe the state of the art of computer RPGs and some academic research that focus on the use of social relationships in autonomous characters. Then we present a description of our model followed by a description of the game that was developed as a case study. Afterwards, we present the study we performed to assess the effects of the model in players' gaming experience and finish we some conclusions.

2 State of the Art

In order to see what we can do to improve social behavior of NPCs in RPGs, we first took a look at how NPCs are portrayed in today's RPGs. Artificial Intelligence (AI) of the NPCs in games is usually implemented with the use of state machines, since it allows for very controlled behavior. However, by using such pre-defined behaviors, in order to attain a rich social behavior from the NPCs, there has to be an extensive authoring. In order to explore a more sophisticated way for the creation of social behavior, we looked at the academic research.

The next two sections show some of our findings on both games and academic community. Lastly, we will present a survey to RPG players in order to get their opinion on the current state of RPGs.

2.1 Role-Playing Games

In order to see how we can improve RPGs we analyzed a few in order to see their flaws: Neverwinter Nights, Knights of the Old Republic II, World of Warcraft, Dungeon Siege II, Fable, and The Elder Scrolls IV: Oblivion. In all of them the NPC behavior is almost the same; the player will engage with NPCs that will provide him with quests that he may or not decide to resolve. The NPCs that supply these quests to the player are usually frozen in time waiting for the player to engage them. If a player accepts the quest, s/he continues with her/his journey, and after resolving the quest returns to the NPC, which will probably still be stuck in time, to receive some kind of reward. NPCs do not have a rich social behavior.

However, game developers are starting to take a closer look at NPCs' AI. Two examples of recent games that try to create richer NPCs are Fable [8] and The Elder Scrolls IV: Oblivion [3]. These RPGs have some features that make them standout from the rest, like some individual believability (e.g. NPCs have jobs and houses) and some social interaction (e.g. NPCs simulate conversation between each other, in Oblivion the player can even hear these conversations, although they clearly sound like tips for her/him).

Nevertheless, even these RPGs have room for improvement. Although there are social interactions, these are not meaningful. For example, in Oblivion the player can enter the house of a NPC and wake her/him up to engage in conversation, and the NPC will simply talk back, ignoring the fact that a stranger just broke-in in the middle of the night. Another situation that we encounter is that we can steal from a NPC and s/he will start calling the police, but if we engage in conversation with her/him, s/he will talk back like nothing happened.

2.2 Academic Community

AI in games mostly feel like finite state machines, if something happens the NPC jumps into a given state. However, in order to have more interesting emergent behavior, this approach is not effective. We decided to explore the academic research and see what ideas regarding social behavior were being explored on autonomous agents, and how they could adjust to the NPCs.

Avatar Arena [15] is a system in which a task is delegated to an autonomous character (i.e. agent). The task in this case is a negotiation between the avatars (i.e. agents that are virtual representations of the users). The idea was to create a system that resembled a human-human negotiation in which not only rational arguments but also the social context and the personalities of the negotiating parties need to be taken into account.

The authors focused on how social relationships could impact an avatar's behavior. To model social relationships between the avatars, they based in socio-physiological theories of cognitive consistency, specifically in the Heider's Balance Theory [6] and Festinger's Theory of Cognitive Dissonance [5].

To model the avatars' personality they based on the Five-Factor Model (FFM) [9] implementing three of its dimensions: Extraversion and Neuroticism because they can be found on most of the theories, and Agreeableness because it is important when dealing with social relationships.

PsychSim is a multiagent-based simulation tool for modeling interactions and influence among groups or individuals [14]. It is based on the fact that the message exchanged between two agents, like in human communication, depends not only on its content but also in the way we see the communicator. Therefore, our actions take into consideration our view of the communicator and also the reaction we expect from her/him.

This means that PsychSim revolves around the notion of Theory of Mind (ToM). Each agent has her/his own decision-theoretic model of the world including her/his beliefs about the environment and her/his beliefs about the other agents' beliefs.

Social Role Awareness was the work of Prendinger and Ishizuka [13], in which they propose that animated agents should have a notion of social roles when interacting with other agents, and adjust their behavior accordingly.

The goal of the authors' work was to create autonomous agents that would be much more interesting conversational partners for language conversation training.

These agents have a standard theory of reasoning about emotions, the OCC model [11] and a simple model of personality that consists of two dimensions from the FFM model, Extraversion and Agreeableness, which according to André et al. [1] are essential for social interaction.

The core of the authors' work is basically a social filter. Before choosing a particular action (e.g. express an emotional state), it will go by the social filter and determine the agent's behavior based on her/his personality and social role.

The μ-SIC System [10] is a social model that consists of an Artificial Neural Network (ANN), which will be responsible for determining which social interactions the character should engage with another.

The μ-SIC System uses three psychological models: the Eysenck personality model [4] the Lang mood model [7], and the relationship model which has its psychological basis on the work of Wish et al. [16].

One thing to note is that this work is the first one that we encounter that focuses on games. Usually academic research is more focused on few agents interacting and try to make the interaction as close to the human interaction as they can, but this one focused on many characters like we find in games, making concessions in order to maintain a responsive real-time experience.

Prada [12] has conducted some research focusing on the problem of engaging users in a group of agents that share a common collaborative task. The believability of the interactions was a key issue, thus, he developed the Synthetic Group Dynamics Model (SDG Model) to increase the believability of the group interactions. The construction of this model was supported on several theories and results of studies of human group dynamics found on social sciences. It defines the knowledge that each individual should build about the others and about the group it belongs, and how this knowledge drives their interactions in the group. One important feature of the SGD Model is that it defines the social relations of the group's members in two different dimensions: (1) the structure of power and (2) the structure of interpersonal attraction. The dynamics of the group is supported on a set of categories that identify the possible interactions that can occur in the group. These interactions are categorized in two main classes, those that are related to the task and those that are related to the socio-emotional relations.

2.3 RPG Gamer Survey

To search for evidence that players, indeed, expect for better social behavior from NPCs, a pilot study was performed. This study consisted in a questionnaire that was published in the Internet and disseminated in several gamers' forums[1]. The questionnaire was removed one month and a half later, with 250 complete answers collected.

[1] Forums from: BioWare Corp, World of Warcraft Europe, GameDev-PT, GameSpot, GamSpy, Lionhead Studios, Mega Score, Obsidian Entertainment, Planet Dungeon Siege, The Elder Scrolls.

The gamers' demographic was 37% between 12 and 18 years old, 39% between 19 and 25, 9% between 26 and 30, 10% between 31 and 40, and 3% over 40 years old. Of all the gamers, 93% were males, and 41% played pen-and-paper RPGs.

One of the aims of the study was to find if there was any significant difference between what players expect from the NPCs and what they found on the RPGs they play. Thus, players were asked to rank (from 1 to 4), regarding what they desire (i.e. how important they feel they are for the gaming experience) and what they found int the games they have played, some features that we believe to be relevant for NPCs' believability: personality, emotions, pro-activeness, individual memory, ability to form social relationships with the player, ability to form social relationships with other NPCs.

When comparing the answers, there was a significant difference between what players desire and what players found regarding all features: personality ($T=1556.50$, $p<0.001$, $r=-0.44$), emotions ($T=1629$, $p<0.001$, $r=-0.36$), pro-activeness ($T=1519.50$, $p<0.001$, $r=-0.40$), individual memory ($T=1639$, $p<0.001$, $r=0.39$), relationships with the player ($T=1346$, $p<0.001$, $r=-0.43$) and relationships with other NPCs ($T=1548$, $p<0.001$, $r=-0.43$).

From the results above, we can see that there is still a lot to improve regarding NPCs believability. In addition, the difference seems to be higher ($r > 0.4$) relatively to personality, and the ability to establish relationships with both the player and other NPCs. This supports our proposal of implementing social relationships in order to enhance the social behavior of the NPCs.

Looking at the correlations between some of the questionnaire's items, we found some interesting results. Players that played pen-and-paper RPGs seem to identify more flaws in computer RPGs regarding characters' personality ($N=250$, $r=0.14$, $p<0.05$) and pro-activeness ($N=250$, $r=0.14$, $p<0.05$). Also, players that play more often expected more from characters' pro-activeness ($N=250$, $r=0.16$, $p<0.01$) and believe more on the importance of improving social behavior of NPCs to improve the role-play experience ($N=250$, $r=0.13$, $p<0.05$). Finally, when asked if they believed that a richer social behavior would improve their gaming experience, 94% answered positively.

3 Conceptual Model

As we could see in the previous section, there is still much to improve in order to attain believable social behavior of NPCs in computer games. Not just our analysis of games tells us this, but also gamers say that they would prefer NPCs with richer social behavior. As mentioned before, we propose to improve the social behavior of NPCs by introducing a mechanism of social relationships between them. This mechanism is based on a concept model that defines for each NPC: Personality, Relationships with others, and Theory of Mind (ToM).

In Fig. 1 we can see the representation of a character's mind, which contains the concepts of personality, relationships and ToM, and how they are connected.

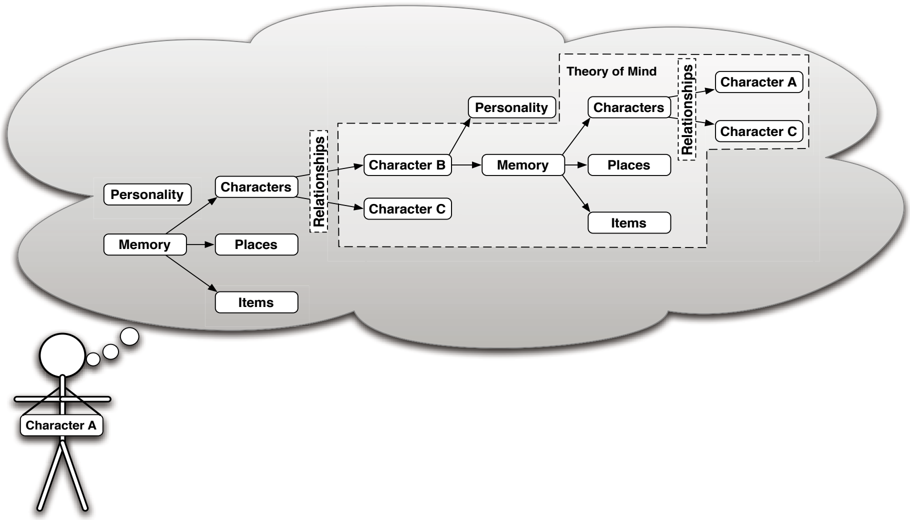


Fig. 1. Mind of Character A

In the following sections we will explain all these concepts and why we believe that all of them are necessary to improve social behavior.

3.1 Relationships

We consider relationships the most important concept because it represents our social connection with each person, it influences how we interact with each other, and without them there is no real social behavior. In RPGs we only have a glimpse of relationships in scripted events, like cut-scenes, which, besides being pre-defined, are scattered along the story and do not create a continuous social behavior.

Some of the recent RPGs try to give a notion of relationships. One common case is NPCs living in the same house, although we do not see behavior such as the man, which we assume that is the husband, kiss the woman before going out.

Our relationship system is based on Heider's Theory of Balance [6], in which relationships are a ternary association between two persons and a concept (see Fig. 2).

For the representation of relationship we consider a simplification of the relationship model proposed by Wish et al. [16], which may be extended in a future

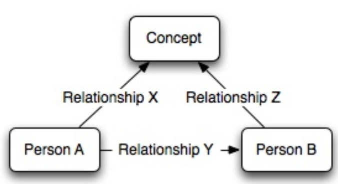


Fig. 2. Heider's Balance Theory

version. We only consider one dimension, Likes/Dislikes axis, since we considered it to be one of the most expressive and easily understandable dimensions.

A high value in this axis can be easily understood as a positive relationship between two NPCs, and the opposite for a low value. We acknowledge that only this axis is not sufficient to accurately map the gamut of relationships, but our focus at this moment is not to accurately map relationships but instead to see if a minimum of relationships will affect the gaming experience of RPG players.

3.2 Personality

In real life, not everyone acts or reacts the same way. For example, if a bully starts beating up a kid in school, it is not sure that everyone will back off or stay and see the fight. Some might also engage in the fight, others might go and find help. Furthermore, not everyone will see the event the same way, some will stay on the bully's side, some will stay on the victim's side, and others will not even care about the incident. Of course there are other factors that influence where we stand, but our personality is at the core of how we might react in a given situation.

Also in games, personality is one of the main differences between the primary (e.g. plot/quest related) and secondary NPCs. It is not common to find a primary NPC's personality in another NPC, but the personalities of secondary characters appear quite often replicated (e.g. all the merchants have the same personality).

We based our concept of personality on a simplified version of the FFM [9], only considering one of its dimensions, Agreeableness. We chose only this dimension for the same reasons that we chose only one dimension to represent relationships, we want a dimension that is easy to understand and measure.

Again, we do not consider that only this dimension is suitable to represent personality, and we believe that it would have to at least include the dimensions present in Eysenck model [4], Extraversion and Neuroticism, since there seems to be a consensus that they are fundamental in implementing relationships [1] [10] [15].

3.3 Theory of Mind

A ToM means that characters have a mental representation about others' beliefs.

In our conceptual model, this means that NPCs have beliefs concerning the other characters' beliefs, specifically their relationships with others. For example, John may believe that Peter likes Mary.

ToM is necessary since it is important for NPCs to be aware of each other, and only then can they have interesting emergent social behavior.

3.4 Behavior Generation

Our model determines two things: (1) with whom the NPC will interact and (2) the quantity of information transmitted in conversations.

The NPC decides with whom s/he will interact based on the relationship with that character, giving preference to characters that s/he likes, and try to avoid characters s/he dislikes. In each conversation, the relationships will influence the quantity of information transmitted (i.e. if they like the character they are talking to and/or the object of the conversation they will give out more information, and the opposite if

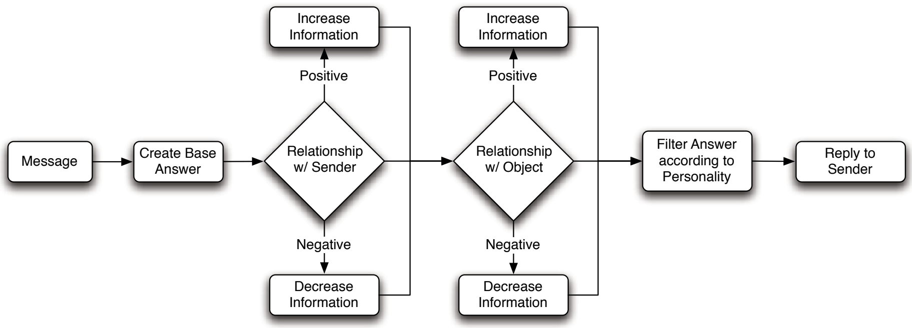


Fig. 3. How the model influences conversations

they dislike), and the personality will influence how the answers are transmitted (i.e. the text, a friendly NPC will be more polite, see Fig. 3).

4 Case Study: Zion's Market

In order to test our proposal we developed a small game that takes place in a market, which is a very common scenario of RPGs.

The player had a malfunction in her/his shuttle crashing on planet Zion, and s/he has to get some items to repair it. In order to complete this task, the player has to find a mechanic to tell her/him what items s/he needs, buy the items from the merchants, and then hire the mechanic to repair her/his shuttle. All these tasks are done through conversations with each of the NPCs (Fig. 4 shows the player having a conversation with an unfriendly merchant).

The game has six NPCs, three are merchants, one is a mechanic, and the other two are customers just like the user. The merchants and the mechanic are standing in their own stands waiting for customers. The two customer NPCs have the ability to enter/leave the scenario (through their shuttles) and trade items with the merchants.

4.1 Sample Interaction

In Fig. 4 we see an example of a conversation between the player and a merchant called John. When the player engaged John, s/he asked him "Do you know who has Energy Cells". In the top part of the figure we can see John's answer.

The first thing we notice is John's personality. By the way he talks we can see that he is rude, hence is personality has a low agreeableness value. The second thing shown in the dialog is that John mentions that he thinks that there is another merchant that sells Energy Cells, a merchant called Jeff. Finally, we can see that John does not like Jeff, since he hints the player not to go see him because "He is really not that nice".

Next the player went to Jade, a merchant that has a stand next to John. She has a neutral personality, not being rude as John, nor being very friendly. When the player asks the same question that s/he asked John, Jade will have a whole different answer. She answers that both John and Jeff sell energy cells, however, because she likes Jeff, she will try to persuade the player to go buy from him.

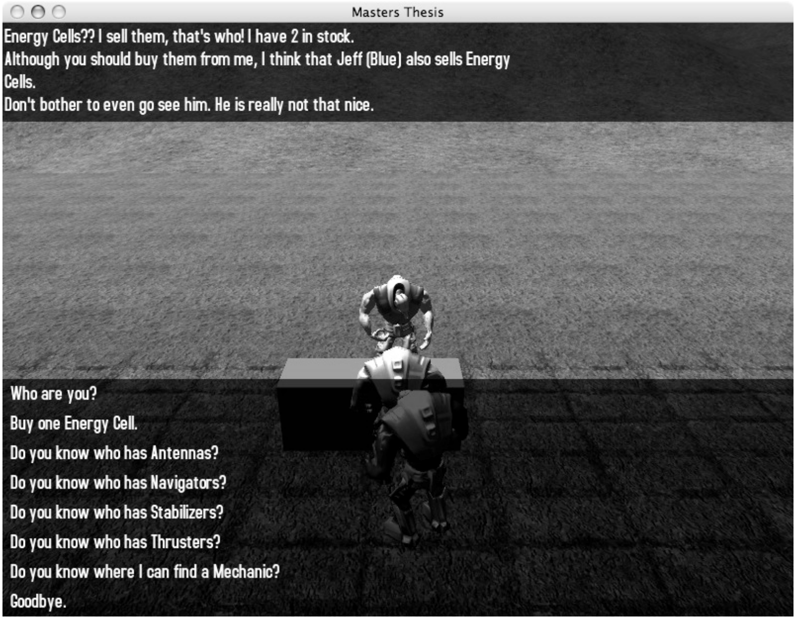


Fig. 4. Screenshot of Zion's Market

5 User Evaluation

This section describes the experiment that we conducted in order to test if social relationships could create a richer social behavior in NPCs, which would lead to an improved gaming experience. This experiment was conducted with the game Zion's Market presented in the last section.

The evaluation was done based on a questionnaire that was given to the users with some questions about the existence and identification of the personalities and the relationships of the NPCs. The logs created from the interaction that the player had with the NPCs (e.g. number of conversations with each one, number of items bought from each merchant) were also recorded.

Each player played the game twice: one time with the model here proposed, and the other without. The players were not told the differences between both versions.

5.1 Method

The evaluation was done with a total of twenty players. There was no time limit, the game ended when the player finished the objective of the game (i.e. hired a mechanic after buying all the items needed to repair the shuttle). The process was divided into 4 parts:

1. Players were given a version of the scenario to play;
2. Players were given a questionnaire about the version they played;
3. Players were then given the other version of the scenario to play;
4. Finally, players answered the same questionnaire about the second version they played.

The whole process took between 20 to 40 minutes with each player.

The version with which the players would start with was evenly distributed, with ten of them starting with version 1, which does not contain social relationships, and the other ten starting with version 2, which contains social relationships, having already played the game and answered the questionnaire when they played for the second time.

The limitations of version 1 were that the characters were not aware of each other, they did not know each other (i.e. did not have memory/ToM of other NPCs), and hence they did not have relationships with each other.

Before the players started the experiment, the game's objective was explained to them. The players were not told which of the versions had social relationships.

The questionnaire had a few questions about personal information (e.g. gender, age), NPCs' personalities (if they thought that the NPCs had different personalities and what kind of personality each one had) and relationships (if they thought that the NPCs had relationships between them, and what kind of relationship was), and a question for them to rate the gaming experience of each scenario. Finally, they were asked what version they preferred.

5.2 Results

This section presents the analysis of the data retrieved from the questionnaire and from the logs of each version of the game that the players played.

We identified that there was a big difference between the ratings that were given to each version ($Z=3.207$, $p=0.001$), with version 2 (i.e. with relationships) receiving a mean rating of 3.45 and version 1 (i.e. without relationships) receiving 2.85.

The questionnaire showed that 80% of all players preferred version 2. Of the players that play RPGs (55% of the all players) 82% of them also preferred version 2, and hence given a difference so little between players that play different types of games and players that play RPGs, one could assume that the introduction of social relationships could improve the gaming of other types of games. One interesting fact is that 100% of the players that played version 1 first, preferred version 2 ($N=20$, $r=0.5$, $p<0.05$).

We analyzed if there was a significant difference between both versions. The results show that there was a significant difference regarding the existence of relationships between the versions ($Z=3.464$, $p=0.001$), which we were expecting since it is the differentiating factor between the two versions. In version 1, 40% of the players answered that the NPCs had social relationships, although they would generally not remember which NPCs had them, nor what kind of relationships they were; approximately 93%, when asked what kind of relationship was, answered that they did not know. In version 2, 100% of the users answered that the NPCs had social relationships between them.

There was a big difference between the number of relationships they identified in the merchants and the mechanic, which makes sense given the fact that the users interacted more with the merchants (a mean of 1.59 conversations with each) and the mechanic (a mean of 2.25 conversations) than with the customers (a mean of 0.82 conversations with each).

It is also worth mentioning that the number of conversations between the player and the merchants is a bit higher in version 2, with 1.65 conversations with each merchant versus 1.53 in version 1. The number of messages exchanged between them in conversations is really much higher in version 2 (9.67 with each NPC, versus 6.33 in version 1), as it was also expected since the players can ask information about the other NPCs.

We also found some interesting results, like the users remembering better the NPCs that had the personalities in the extremes of the personality axis (friendly 65% and unfriendly 77.5%).

6 Conclusions

Clearly with the maturing of the computer games' market, game studios start doing more elaborate games. Like with the progression of graphical quality, it is only natural that developers start to focus on areas yet to release their full potential, like in the case of NPC AI.

The purpose of this work was to test the hypothesis that if we enrich the social behavior of the NPC by implementing social relationships, the player's gaming experience can be improved.

To test this hypothesis, a conceptual model was developed taking into account the concepts of relationship, personality, and theory of mind. An implementation of this model was used to develop a game that was evaluated by players.

The players played both version of the game, one with social relationships, and the other without, and in the end 80% of them preferred the version with social relationships.

Although we believe that this could be the basis for mode believable social behavior, we also believe that much is still to be done. The next step would be to expand the concept model to more accurate definitions of relationship and personality, followed by the inclusion of social roles, which we believe that are also very important when representing rich social behavior.

Acknowledgments

The work presented in this paper was supported by Mind RACES (Contract 511931) and eCircus (Contract no. IST-4-027656-STP) projects carried out with the provision of the European Community in the Framework VI Programme.

References

1. André, E., Klesen, M., Gebhard, P., Allen, S., Rist, T.: Integrating models of personality and emotions into lifelike characters. In: International Workshop of Affect in Interactions. Towards a New Generation of Interfaces (1999)
2. BBC News: US Video Games Sales Hit Record, `http://news.bbc.co.uk/1/hi/business/7195511.stm`

3. Bethesda Softworks: The Elder Scrolls IV: Oblivion, `http://www.elderscrolls.com`
4. Eysenck, H.J., Rachman, S.: The Causes and Cures of Neuroses. Routledge/Kegan Paul, London (1965)
5. Festinger, L.: A Theory of Cognitive Dissonance (1957)
6. Heider, F.: The Psychology of Interpersonal Relations (1958)
7. Lang, P.J.: The Emotion Probe: Studies of motivation and attention. A study in the Neuroscience of Love and Hate (1995)
8. Lionhead Studios: Fable, `http://fable.lionhead.com`
9. McCrae, R.R., John, O.: An introduction to the five-factor model and its applications. Journal of Personality 60, 175–215 (1992)
10. Namee, B.M., Cunningham, P.: The Driven Simulation of Socially Interactive Agents (2002)
11. Ortony, A., Clore, G.L., Collins, A.: A Cognitive Structure of Emotions. Cambridge University Press, Cambridge (1988)
12. Prada, R., Paiva, A.: Believable Groups of Synthetic Characters. In: 4th International Joint Conference on Autonomous Agents and Multi Agent Systems. ACM Press, Utrecht (2005)
13. Prendinger, H., Ishizuka, M.: Social Role Awareness in Animated Agents. In: 5th International Conference on Autonomous Agents, pp. 270–277. Montréal (2001)
14. Pynadath, D.V., Marsella, S.C.: PsychSim: Modeling Theory of Mind with Decision-Theoretic Agents. In: International Joint Conference on Artificial Intelligence, pp. 1181–1186 (2005)
15. Rist, T., Schmitt, M.: Avatar Arena: an attempt to apply socio-physicological concepts of cognitive consistency in avatar-avatar negociation scenarios. In: AISB 2002 Symposium on Animated Expressive Characters for Social Interactions, London, pp. 79–84 (2002)
16. Wish, M., Deutsch, M., Kaplan, S.: Perceived Dimensions of Interpersonal Relations. Journal of Personality and Social Psychology 33(6) (1976)

A Surround Display Warp-Mesh Utility to Enhance Player Engagement

Charles Henden[1], Erik Champion[2], Ralf Muhlberger[3], and Jeffrey Jacobson[4]

[1] School of Information Technology and Electrical Engineering,
The University of Queensland, St Lucia QLD 4072, Australia
charles@henden.com.au
[2] College of Fine Arts, University of New South Wales
Sydney NSW 2052, Australia
erikchampion@cofa.unsw.edu.au
[3] School of Information Technology and Electrical Engineering,
The University of Queensland, St Lucia QLD 4072, Australia
ralf@itee.uq.edu.au
[4] PublicVR, 333 Lamartine St., Jamaica Plain, MA, USA, 02130
jeff@publicvr.org

Abstract. Surround displays are used in simulation, training, and other applications based on virtual worlds. A wide-view display engages the viewer's peripheral vision, providing a more accurate view of the virtual world and therefore a heightened sense of immersion. However, most commercially available surround displays are expensive and complex. We developed a low-cost alternative, which uses a standard digital projector, a hemispherical mirror, and any roughly spherical or cylindrical screen. The software can handle irregular surfaces and will be open-sourced in the next release of the CaveUT/VRGL freeware. We also conducted a pilot study comparing game play in our prototype and game play with a standard desktop monitor. Players using the surround display reported significantly shorter (P = 0.0051) perceived duration of time during play. Reduced awareness of the passage of time during game play was positively correlated with greater engagement and enjoyment.

Keywords: Virtual Environments, Virtual Reality, Immersion, Spherical Mirror Projection, Dome Projection, CaveUT, VRGL, Surround Gaming, Virtual Reality, Warp-Mesh, Texture-Mesh.

1 Introduction

Three-dimensional virtual environments for educational and entertainment purposes are most commonly viewed via a small, flat display. Yet a traditional flat display cannot express the dimensions of a virtual scene relative to the user. Most importantly, it cannot offer the spatial cues provided by engaging a user's peripheral vision and head movement. Engaging peripheral vision is a significant component of participating in a virtual world, as is experiencing the environment in its correct scale [1]. Allowing a user to interact in the same space as a virtual object is a powerful method of providing immersion, or the sense of 'being there' [2].

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 46–56, 2008.

Believable virtual spaces provide new forms of collaboration and expression. A viewer can become mentally absorbed in an artificial environment by cues such as peripheral engagement, binocular disparity, convergence (the degree to which the eyes rotate toward the centre of interest), perspective projection [1], and the act of controlling the animation presented in the material.

Fully featured virtual reality systems, such as the CAVE [1], provide nearly all of these cues, while others provide only some. For example, in ‘Fish Tank VR’ a single computer monitor produces a stereographic image while a head-tracking system adjusts the perspective effects based on user motion. Other systems achieve a degree of sensory immersion by simply filling most or all of the user’s view with a “window” to the virtual environment. We call these “surround displays,” with the new all-digital planetarium displays being the best example.

Fig. 1. Drawing of our prototype surround display, 110° horizontal FOV

We[1] constructed our own low-cost surround display (Figure 1) and used it to study the effect of visual immersion on user engagement. In this paper we describe how the user-defined warp-mesh utility works and how to use it. We present this design as a low-cost alternative to more expensive immersive displays.

2 Research Problem

Both CAVEs and Head Mounted Displays (HMDs) are typically expensive and fragile and present difficulties for application designers, which reduces their widespread

[1] Charles Henden did most of the work, producing the warp-mesh code and conducting the user engagement study as his Master’s Thesis work, supervised by Erik Champion and Ralf Muhlberger. Jeffrey Jacobson’s VRGL and CaveUT software provided the software platform.

use. That is why three-dimensional virtual environments for educational and entertainment purposes are most commonly viewed via a small, flat display even though a traditional flat display cannot express the dimensions of a virtual scene relative to the user.

Although recent research suggests that large high-resolution displays improve the user experience [3] and ongoing research is testing planetaria with game-based content [4][5], projection via virtual objects [6], and automatic projector calibration [7], we know of no research into user-modified warping mesh for home entertainment. Warping calibration typically involves an expert who sets up and calibrates an application for a specific environment. For home entertainment, on the other hand, software needs to be flexible, easy to use, affordable, and free of any requirement for programming knowledge.

Modifying an existing game eliminates the need for specialized software to create a multiple-projector virtual reality system [8]. Even so, the overhead of acquiring and positioning a minimum of two digital video projectors and dedicated client computers constrains the average classroom or gamer from utilizing such a system. The shortcomings of both head-mounted displays and CAVEs indicate the need for an inexpensive surround-projection method. Recent developments in projection methods specific to planetarium installations may provide a way to bring the advantages of a surround display [9] to a wider audience. In our surround display, the virtual environment is projected onto a spherical mirror to create a much larger projection than previously possible, and source image correction is provided to deal with distortions created by the mirror and projection surface.

However, there is still a limiting factor to bringing immersive projected displays to a wider audience: the projection surface. As a low-cost alternative, we use a single projector and spherical mirror along with correctional software to project on all manner of surfaces, from professional domes to temporary cylindrical or polyhedral spaces and even the corner of an ordinary room.

2.1 Bourke's Spherical Projection

Recently, all-digital projection systems have become popular for planetaria and similar venues which feature a visually-immersive dome-shaped projection screen. Many of them use a single projector with a fisheye lens to spread the image across the entire screen, often providing a view as wide as 180°. Unfortunately, these lenses are very expensive. As a low-cost alternative, Bourke [5][10][11][12][13] found a way to use a convex mirror as a low-cost substitute for the fisheye lens to distribute the image.

In that system, source material must be pre-distorted or modified according to an analytical mapping which represents the dome and mirror. One example projection method is to produce a pre-distorted image which is the inverse of the distortions the mirror and the screen would introduce. The end result is an image which looks correct on the dome. Traditionally, all corrective image manipulation has occurred prior to the projection of the source material. However, interactive environments require that all changes to the rendering output occur in real time.

For either case, static or dynamic, Bourke describes two basic methods to specify the transformations needed. People have been interested in using curved mirrors for projection for thousands of years [14], so our understanding of the required image transformations rests on a considerable body of knowledge.

The first step is to build a mathematical model of the projection environment by (theoretically) tracing a single ray for each pixel from the projector lens to the spherical mirror and then onto the projection dome. Software can use the general formula for calculating each ray to project faithfully on the dome. This is important for software which produces images in real-time, such as a game engine. Simulating the geometry of a new projection surface is the most precise way of correcting for projection distortions. However, the overhead and expertise required for each projection surface is considerable, and the need to keep the surface as a simple shape (a partial sphere or cylinder, usually) is acute.

Bourke suggests an alternative, which is to allow the user to interactively specify the needed transformations. This splits the task of creating an immersive virtual environment through interactive mesh creation into two separate components. The first is to create the "best-fit" corrective mesh via a standalone application, referred to here as the Warp-Mesh. The second part is to modify the rendering output of a pre-existing virtual environment application according to the warping mesh created by the warp-mesh utility.

This approach can handle many irregular projection surfaces and is the most flexible way of dealing with ad-hoc projection environments. We have adopted this approach for our prototype display. For us, the final image produced was of somewhat lesser quality than with an analytic model, but further research and development may correct that disadvantage. In any case, it is a flexible and low-cost method which can bring surround displays to a wider audience of users and developers.

2.2 CaveUT and VRGL

CaveUT is an open-source freeware project that uses the game Unreal Tournament 2004 to produce immersive projection-based virtual reality. This software operates using the specification of a regular CAVE, allowing each projector to recreate part of a contiguous view of the environment [11][15]. Virtual Reality Graphics Library (VRGL) is a companion project to facilitate off-axis projection in any application. It utilizes the OpenGL® rendering library by re-interpreting the commands created by an application to render a virtual scene. VRGL can reproduce the resulting image as a part of a multi-screen immersive display.

VRGL also supports features useful in the setup of immersive displays: spherical distortion correction, edge blending and color correction. To implement our prototype display, we added to VRGL the capability of loading warp-mesh format files, generated by our interactive modeler. Due to the integration between Unreal Tournament and the CaveUT system, VRGL had been used extensively in conjunction with Unreal Tournament to produce a variety of immersive displays. For this reason, it was chosen for this pilot study.

3 Design

We developed an application to generate and edit a best-fit model for any projection environment suggested by Bourke. To evaluate this approach, we built a low-cost prototype display, which included a projection screen, a projector, a hemispherical mirror, and software. The software was an extension to Jacobson's VRGL graphics library as used in the CaveUT project [16] to support spherical mirror projection.

The specification for the mesh-warping file is very similar to the original format specified by Bourke [10]. The warping is performed in the x, y, z coordinates, in the u, v coordinates, or both. The spatial coordinates of x, y and z allow correction to be made for the surface on which the virtual environment will be projected. The texture coordinates of u and v define the warping for the convex mirror that is to be utilized. The file format specifies a number of parameters:

- The first line contains the mesh type of 1 for polar and 2 for rectangular as described in Bourke's research.
- The second line contains the mesh dimensions, which usually correspond to a multiple of the aspect ratio of the source material.
- Subsequent lines define the nodes of the mesh, containing the following 8 values, each expressed as a floating point number delimited with a space character:
- Spatial coordinates for x, y and z in normalized OpenGL space. Upon mesh generation, all z values will be identical.
- Texture coordinates u and v, ranging from +/-1.0f.
- A value for r, g and b color value to be applied to the mesh. These values are used as a multiplicative intensity for per-color, gamma-corrected edge blending. An additional use for these values is correction for brightness and material differences in compound projection surfaces. Values outside the 0.0f to 1.0f range will be clamped to their minimum or maximum.

A utility to create these warping meshes must be easy to use and must incorporate the majority of common functions required for a variety of projection surfaces. For example, to facilitate projection into the corner of a room, the mesh must be able to be deformed about the centre, stretching the scene locally to avoid information loss at the crease of the wedge. The size, shape and scale of the mesh must be completely editable in order to provide the flexibility to adapt to any projection surface.

We built a modification to VRGL to allow warping, and a GUI (Figure 2) to allow a typical user to change the nodes of the warping mesh by eye. The software tool to read, write, and generate warp-mesh files was written in C++ and uses the OpenGL graphics library to display the mesh while editing and to preview the results of the completed warping. The user (editor) does not need to be a programmer or be highly computer-literate in order to use the software. We then created a curved surface on which to project (Figure 3).

The experiment was conducted using a typical low-end gaming PC with a GeForce™ 6200 video card used for rendering the virtual environment via OpenGL 1.5. The computer was connected simultaneously to a projector for the surround display and to a traditional computer monitor. The traditional display consisted of a 17-inch LCD monitor with a resolution of 1024 x 768 pixels. The projector was a LCD model commonly used for data display, matching the resolution of the LCD monitor.

The surround display's projection surface consisted of a set of two curved display stands covered in white blackout cloth. The surface was 2.4 meters wide, situated 1.2 meters from the participant, yielding an effective field of view of 110 degrees (Figure 3). The curved display stands were constructed especially for the study from low-cost,

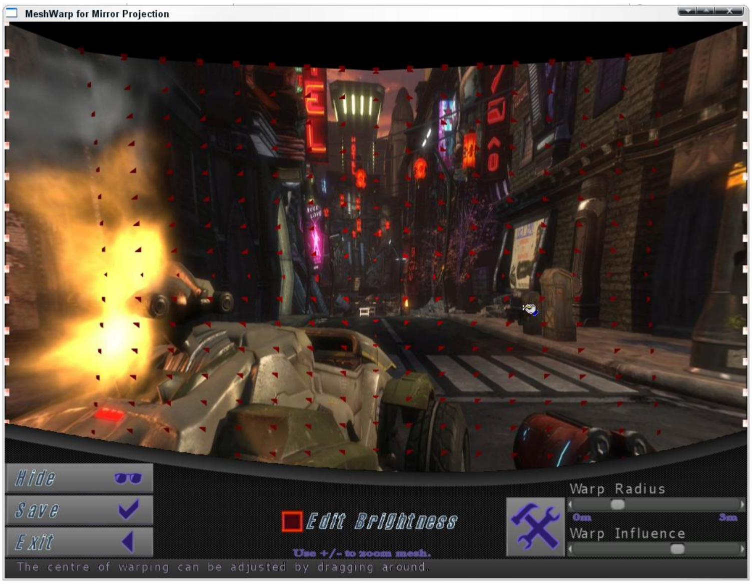


Fig. 2. The GUI allows any user to change the warping mesh graphically

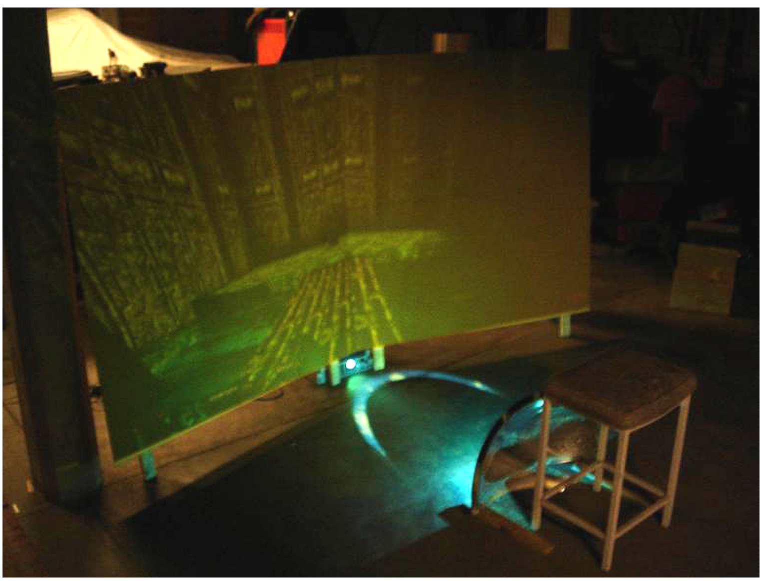


Fig. 3. Experimental display system with curved mirror

readily available medium-density fiberboard and tubular steel framing. The surround screen and LCD monitor were located next to each other to facilitate a quick change between the two displays.

4 Experimental Study and Evaluation

4.1 Hypothesis

For this study, we proposed that the immersive effect of the surround display would be greater than that of a flat display because a surround display engages the participant's peripheral vision and covers a much larger field of view than a traditional display. We then hypothesized that the display that produced the largest differential between actual time taken to complete the task and the subject's perception of the time taken would be categorized as producing the most immersion. Also, and conversely, we hypothesized that the display for which subjects would report more time taken to complete the task (compared to the actual amount of time taken) would indicate boredom and a lack of relaxation.

Of primary interest were the quantitative results pertaining to the amount of time participants spent completing each display in contrast to how long they perceived they spent. For many gamers, losing track of time is a positive experience and is one of the main reasons for playing video games [17]. Table 1 shows the differential between actual time and perceived time for both the surround and traditional display for each participant. A negative time shows that the time the participant believed was spent completing the task was longer than the time actually spent. Conversely, a positive time indicates the participant losing track of time and spending longer in the virtual environment than he or she believed.

4.2 Protocol

We recruited twelve participants from the undergraduate computer science laboratories by offering each one a randomly drawn prize for their participation. We asked them about their familiarity and usage of virtual environments, specifically computer and video games. Of particular importance was the number of hours each subject spent playing 3D games per week and what display the subject usually employed.

We randomly assigned each subject to use either a standard desktop monitor or our prototype surround display. Then we showed the subject how to navigate in a virtual environment using the physical controls, answered any questions, and asked the subject to find twenty-five virtual goal objects. This forced the subject to navigate through a majority of the environment, creating a similar experience for everyone. The experimenter used a stopwatch to record the time taken.

After completing the first task, the subject switched to the other display type and began another hunt for twenty-five goal objects in a different virtual environment. It was important for the second task to begin without a break so that the subject would not forget the impression of switching displays. The experimenter also timed this exploration task. Switching the overhead fluorescent lights off during use of the surround display controlled the amount of light in the room.

The two virtual environments had different maps to eliminate a training effect between one task and the other. Randomly assigning the order in which each test subject completed the tasks eliminated ordering effects. Finally, there was probably little if any training effect in the use of the controls because they were fairly straightforward and familiar to our test population.

After finishing the second task, the subjects completed a questionnaire detailing their experience. Most importantly, the questionnaire asked them how long they thought they spent performing each task. It also asked subjects to record which display made them feel like they were "in the game" along with the strengths and weaknesses of each display.

The study attracted participants with a moderate to high level of experience with computer-generated 3D environment and a high frequency of game playing, due to the description of the study published on notices and the nature of the prize – a video game store gift voucher. Although the participants were not pre-selected, the method of recruitment resulted in an all-male sample group. This was mainly due to gender distribution in the undergraduate computer science courses and, to a lesser extent, to the tendency of males to play video games much more often than females [17]. There were no explicit prerequisites to participation in the study and all participants were advised on the possibility of motion sickness induced by the movement of the virtual environment.

4.3 Questionnaire

In order to correct the results from each participant, the pilot study questionnaire was designed in the following format. Names and gender of participants and their answers to the following questions were recorded.

1. On average, how many hours a week would you spend playing computer or video games that involved a three dimensional environment? 1-2, 2-4, 4-8, 8 or more
2. What games are you playing regularly at the moment (please list)?
3. What kind of display do you use when playing games? CRT Monitor, LCD Monitor, Projector, Other (please specify).
4. What is the approximate size of the display you use when playing games? 15" or smaller, 17", 19", 20" and larger?
5. What input devices do you usually use to control computer and video games? Joystick, Game pad, Keyboard Only, Mouse Only, Keyboard and Mouse, Other (please specify).
6. In what posture are you most comfortable playing video games? Standing, Sitting, Lying down, Other (please specify).
7. Which display method did you prefer to view the game with? LCD Display, Surround Display?
8. What were the strengths (if any) of using the LCD display?
9. What were the weaknesses (if any) of using the LCD display?
10. What were the strengths (if any) of the surround display?
11. What were the weaknesses (if any) of using the surround display?
12. Did either display give you the sensation of being immersed inside the game world? LCD Display, Surround Display, Neither?
13. How long did you think you played for (in minutes)?

4.4 Results

For each subject and for each task, we compared the actual time it took with how long the subject thought it took. If perceived time was less than actual time, we called that a *time loss*, and the opposite a *time gain.* While using the surround display, most participants experienced some degree of time loss. The average subjective time loss was twenty-one seconds, which is eleven percent of the average time taken. In contrast, while using the standard computer monitor, subjects experienced a subjective time gain of one minute and two seconds, which is forty-seven percent of the average time taken.

We did not expect that such a reliable degree of time gain would be recorded for the traditional display. This effect may be due to the familiarity of participants with that display method, with eighty-three percent of participants using a monitor for over two hours per week. Subjects who spent eight or more hours per week playing games which involved a 3D environment recorded the greatest time gain figure for the task using the traditional display. Their time loss for the surround display was not significant.

Table 1. Time loss differential experienced by participants for each display type

Number	**LCD (flat monitor) Actual - Subjective Time**	**Surround Display Actual - Subjective Time**
1	-1.46	2.13
2	-0.14	.1
3	0.02	0.17
4	-1.65	-0.47
5	-1.91	-1.47
6	0.35	1.56
7	-1.47	0.03
8	-2.96	0.35
9	-1.84	0.45
10	-0.09	-0.25
11	-0.34	0.24
12	-0.7	-0.29
	Total -12.19	**Total 2.55**
	Average -1.02	**Average 0.21**

A standard **T-Test** on the data shows a highly significant difference between the time loss/gain experienced by the two groups:

P = 0.005

Table 1 shows the subjective time loss/gain for each student while using each display. When we compare the results from the use of each display, a two-tailed uneven-variances T-Test shows a high degree of certainty that the subjects are reporting significantly different time differentials for the two displays, with P = 0.005, which means there is a 99.5% chance that the observed difference is genuine and not the result of a random variation. The averages and totals show that subjects felt the passage of time was less in the surround display. We believe that the less the player feels the passage of time (time loss) the more that player is concentrating on and enjoying the task [8]. From this we conclude that the surround display provided a more engaging experience.

The small and homogeneous sample size used for this experiment does limit the applicability of our result to other populations. The all-male sample also may have skewed the results, because males are more likely to report losing track of time while playing video games than females do [18]. Participants who spent less than two hours per week playing games that involved a 3D environment showed a time loss figure of ten seconds or less, signifying that immersion through 3D games is perhaps a learned ability or appreciation.

Regardless of whether the ability to become immersed in a 3D environment is heightened through repeated regular use, the disparity of times between the traditional and surround display show that there is difference between the displays in the ability for users to lose track of time. While most of the participants recorded that they preferred using the surround display compared to the traditional display, all participants recorded some level of immersion experienced for the surround display.

Of additional value were the qualitative comments relating to the strengths and weaknesses for each display method. The most prevalent comment recorded under the surround display's strengths was that the projected display felt more real. Ninety-one percent of participants recorded an emotive preference for the projected display with sixty-six percent of comments including phrases similar to "feels like you are in the game" or "felt more real" or including the word "immersive". One participant recorded this response: "Much better feel for the game, Better gaming experience. Player feels like they are in the game, more... exciting?"

In comparison, the most common strength recorded for the traditional display was the clarity and brightness of the display, a very noticeable property especially if the projected display was trialed first. In contrast, the most commonly recorded weaknesses for the projected display related to poor brightness, low image quality and blurriness produced by distortions in the projection surface and sub-par lighting conditions.

5 Conclusion and Future Research

Projected surround display via interactive warp-mesh correction is shown to provide a noticeable level of immersion. Therefore, spherical projection is a new, low-cost display option for immersive virtual environments. Given more time, we would have also evaluated the usability of the interface for gamers to configure and adjust the system to suit the often changing conditions of their environment.

Further possibilities for research into alternative surround display systems include the evaluation of head tracking and stereo virtualization tools in conjunction with a projected surround display as proposed in this document. A control method for surround display systems allowing three degrees of freedom would also be of value. There may also exist a commercial opportunity in creating a flexible and inexpensive surround display for the video games and computer entertainment industry, and covering not just OpenGL but also DirectX® games.

References

[1] Cruz-Neira, C., Sandin, D.J., DeFanti, T.A.: Surround-screen projection-based virtual reality: the design and implementation of the CAVE. In: Proceedings of the 20th annual conference on computer graphics and interactive techniques, pp. 135–142. ACM, New York (1993)

[2] Slater, M., Usoh, M., Steed, A.: Depth of Presence in Virtual Environments. Teleoperators and Virtual Environments 3, 130–144 (1994)

[3] Sabri, A.J., Ball, R.G., Fabian, A., Bhatia, S., North, C.: High-resolution gaming: Interfaces, notifications, and the user experience. Interacting with Computers 19, 151–166 (2007)

[4] Burney, T., Lock, P.: Measuring Game-Play Performance and Perceived Immersion in a Domed Planetarium Projection Environment. In: Ma, L., Rauterberg, M., Nakatsu, R. (eds.) ICEC 2007. LNCS, vol. 4740, pp. 22–27. Springer, Heidelberg (2007)

[5] Bourke, P.: Immersive gaming: a low cost projection environment. In: DIME 2006, Bangkok, Thailand (2006)
[6] Ofek, E., Rappoport, A.: Interactive reflections on curved objects. In: Proceedings of the 25th annual conference on computer graphics and interactive techniques, pp. 333–342. ACM, New York (1998)
[7] Lee, J., Dietz, P., Aminzade, D., Hudson, S.: Automatic Projector Calibration using Embedded Light Sensors. In: Proceedings of the ACM Symposium on User Interface Software and Technology, pp. 123–126. ACM Press, Santa Fe (2004)
[8] Lewis, M., Jacobson, J.: Game Engines In Scientific Research. Communications of the ACM 45, 27–31 (2002)
[9] Demiralp, C., Laidlaw, D.H., Jackson, C., Keefe, D., Zhang, S.: Subjective Usefulness of CAVE and Fish Tank VR Display Systems for a Scientific Visualization Application. In: 14th IEEE Visualization 2003 (VIS 2003), p. 86 (2003)
[10] Bourke, P.: Mesh format for image warping, WASP, Perth, Western Australia, vol. 2008 (2006) (Describes a file format for storing warping meshes)
[11] Bourke, P.: Spherical mirror: a new approach to hemispherical dome projection. In: Wyvill, G., Billinghurst, M., Arnold, D. (eds.) Proceedings of the 3rd international conference on computer graphics and interactive techniques in Australasia and South East Asia, pp. 281–284. ACM Press, Dunedin (2005)
[12] Bourke, P.: FAQ: Spherical mirror projection for hemispherical dome projection. WASP UWA online, Perth Western Australia (2008) (online article)
[13] Bourke, P.: Using a spherical mirror for projection into immersive environments. Graphite (ACM SIGGRAPH), Dunedin, New Zealand, pp. 281–284 (2005)
[14] Mihas, P.: Some historical resources for teaching about curved mirrors. Physics Education 37, 311–317 (2002)
[15] Jacobson, J., Le Renard, M., Lugrin, J. L., Cavazza, M.: The CaveUT System: Immersive Entertainment Based on a Game Engine (2005)
[16] Jacobson, J., Kelley, M., Ellis, S., Seethaller, L.: Immersive Displays for Education Using CaveUT. In: Jacobson, J. (ed.): World Conference on Educational Multimedia, Hypermedia and Telecommunication. AACE, Montreal, Canada (2005)
[17] Cragg, A., Taylor, C., Toombs, B.: Video Games: Research to improve understanding of what players enjoy about video games, and to explain their preferences for particular games. British Board of Film Classification, London, pp. 1–107 (2007)
[18] Wood, R.T.A., Griffiths, M.D., Parke, A.: Experiences of Time Loss among Videogame Players: An Empirical Study. CyberPsychology & Behavior 10, 38–44 (2007)
[19] Mania, K., Chalmers, A.: The effects of levels of immersion on memory and presence in virtual environments: a reality centered approach. Cyberpsychology & behavior: the impact of the Internet, multimedia and virtual reality on behavior and society 4, 247–264 (2001)

Development and Evaluation of a Centaur Robot

Satoshi Tsuda[1], Kuniya Shinozaki[1], and Ryohei Nakatsu[2]

[1] Kwansei Gakuin University, School of Science and Technology
2-1 Gakuen, Sanda, 669-1337 Japan
{amy65823,scbc0052}@kwansei.ac.jp
[2] National University of Singapore, Interactive & Digital Media Institute
Block E3A, #02-04, 7 Engineering Drive 1, Singapore 117574
idmdir@nus.edu.sg

Abstract. Recently various types of robots are being studied and developed, which can be classified into two groups: humanoid type and animal type. Since each group has its own merits and demerits, a new type of robot is expected to emerge with greater strengths and fewer weaknesses. We have proposed a new type of robot called the "Centaur Robot" by merging the concepts of these two types of robots. This robot has a human-like upper body and a four-legged animal-like lower body. Due to this basic architecture, we have found that the robot has several merits, including human-like behaviors and stable walking. This time we have carried out the experiments to evaluate the stability of its waling motion. This paper described the detailed results of the experiments as well as the construction of the software/hardware of the robot.

1 Introduction

It is expected that in the near future various kinds robots would be introduced into our home. By interacting with us using verbal/nonverbal communication functions, they would support us in our everyday life and/or entertain us. Various robots for such purpose are being studied and developed in research institutes and companies that can be classified into two groups: a humanoid robot with two legs [1][2], an animal type robot with four or more legs [3][4][5][6]. Also a humanoid robot can be classified into those with two legs and those with wheels [6]. Each of these types has its own merits. The design of a humanoid robot with two legs is based on humans and can mimic such human motions as walking. Since this robot's behavior resembles human behavior, it might easily be introduced into society. At the same time, however, its walking capability still lacks stability, and it sometimes falls down, restricting its area of activity. Also it has difficulty maintaining its balance on ground that is not flat. On the other hand, the merit of an animal type robot is its four legs, which allow it to walk stably even on uneven ground. Since it can also basically stand on three legs, it can adopt to various ground pattern changes. So far, however, the robot has mainly been developed as a pet to which useful applications have rarely been applied. A humanoid robot with wheels for locomotion, which we call a wheel type robot, can move very smoothly and stably on the ground. It rarely falls down. It can even move on slightly uneven ground. On the other hand, it has no ability to move on stairs, which greatly restricts its area of activity since houses usually contain stairs and other types of height differences.

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 57–66, 2008.

In the previous paper, we have proposed an approach to overcome these problems. We proposed a new type of robot with a human-like upper body and an animal-like lower body that we call a "Centaur Robot [7]." In the previous paper, we described its basic concept and then its detailed software/hardware architectures. Also we pointed out that the robot could achieve stable walking motions. This time we carried out several experiments to quantitatively evaluate the stability of walking. In this paper we first simply introduce the software/hardware construction of the robot. Then we will describe the details of the experiments.

2 Related Works

Recently, especially in Japan, various kinds of robots have been studied and developed, particularly humanoid robots that are expected to support our daily life. For example, HONDA has developed a humanoid robot called ASIMO that has sophisticated walking capability [1]. For animal types of robots, on the other hand, most have been studied and developed as pets instead of supportive robots, including AIBO developed by Sony [3].

Although much research/development continues on humanoid and animal types of robots, little research has integrated these two types for several reasons. One reason is that since there are so many research themes for new functions and improvements for each of these types of robots, researchers have little incentive to concentrate on new types of robots that go beyond humanoid or animal types. Another is that even myths or folktales only contain a few examples of such creatures as centaurs, mermaids, and sphinxes in which humans and animals are integrated. Thus it is rather hard to imagine the functions and application areas that such a new type of robot might have.

Taking these situations into considerations, we have proposed and developed a centaur robot because we believed by integrating two types of robots we could create a new type of robot with advantages over conventional robots [7]. In the previous paper, we indicated that the robot has a capability of stable walking motions. In this paper we will describe the detailed results of the walking stability evaluation.

3 Humanoid Robots

In our work, we are developing a robot that can stably achieve various motions by merging two types of robots: a humanoid and an animal.

There are two approaches for such integration: from the humanoid robot side and from the animal robot side. The former approach tries to realize a four-legged body as well as four-legged walk while maintaining a human-like upper body and achieving human-like motions. On the other hand, the latter approach achieves various human-like motions by adding a human upper body to a four-legged robot. In our study, we chose the former approach and modified the hardware and software of a humanoid robot to realize a centaur robot.

We adopted a humanoid robot developed by Nirvana Technology as a platform robot [8][9][10]. This robot has 22 servo motors that can express various human-like motions. Figure 1 shows its appearance, and Table 1 shows its specifications. Figure 2 illustrates the construction of its hardware. The control board, on which a microprocessor SH2 is attached, is connected to the servo motors, a gyro sensor, acceleration sensors, PC, and a battery. The program on the controller can achieve autonomous robot behaviors. At the same time, we can send commands to the robot by PC.

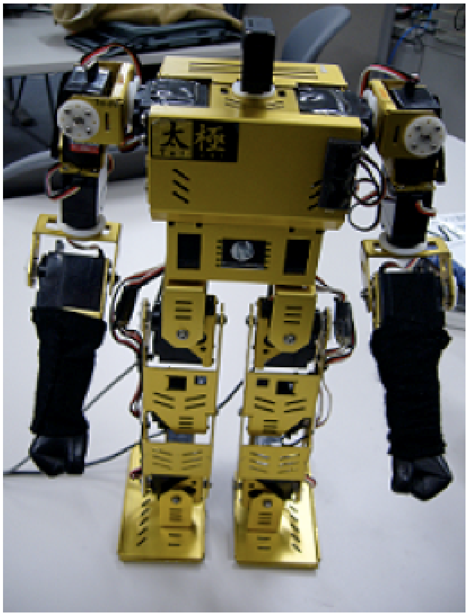


Fig. 1. Humanoid robot

Table 1. Specifications of humanoid robot

Size/Weight	34 cm/1.7 kg
Degree of flexibility	22 (12 legs, 8 arms, 1 waist, 1 head)
CPU	SH2/7047F
Motor	KO PDS-2144, FUTABA S3003, FUTABA S3102, FUTABA S3103
Battery	DC6V

We have carried out various kinds of studied using this prototype robot. We have developed a robot that can achieve Chinese exercise called Tai-chi [8]. Also we have proposed and developed a robot system that can achieve various kinds of dance motions [9][10]. Through these studied we have confirmed that this robot has a stability, flexibility, and sustainability to be used to various kind of applications. Thus we have adopted this robot for our study.

Figure 3 illustrates the software construction. The calculation of the commands necessary to move each motor is carried out each fifteen milliseconds and sent to each servo motor. The instructions to the robot from the PC are first analyzed and based

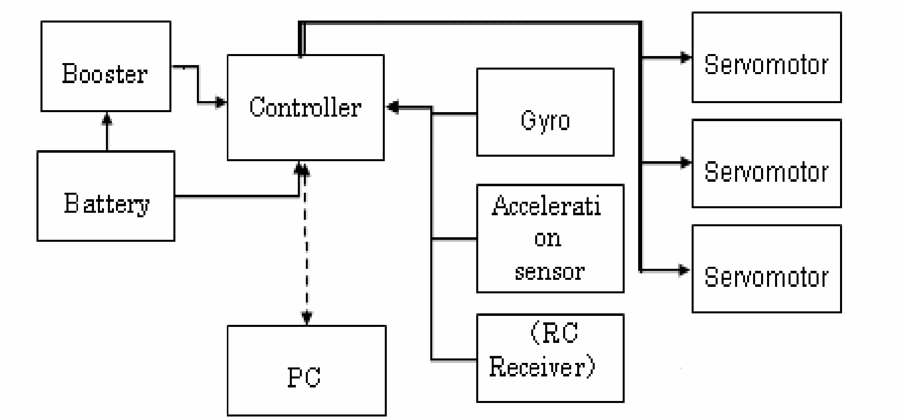


Fig. 2. Hardware construction of humanoid robot

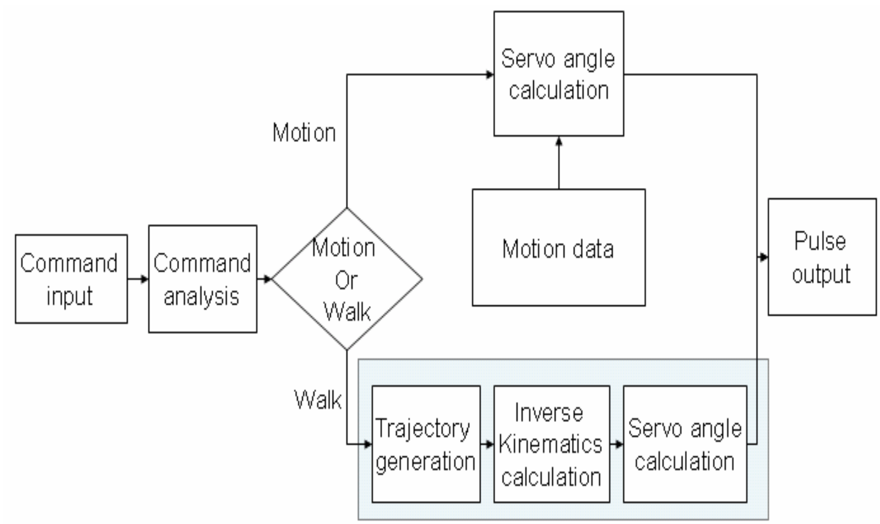


Fig. 3. Software construction of humanoid robot

on results go through one of two processes: one command for walking and other commands for other motions. For other commands, the motion data corresponding to the command is read from memory and the control data for each motor is calculated, and then the control data is sent to each servo motor. On the other hand, if the input command is a command for walking, then the real time calculation of the control data for each servo motor is carried out and sent to each servo motor. Calculation consists of three processes: trajectory generation calculation, inverse kinematics calculation, and servo motor angle calculation. In trajectory generation calculation, the position of each ankle studied by observing human walking motion is calculated every fifteen seconds. Then by inverse kinematics calculation the rotation angle of each foot joint is calculated for the same timing. Based on these calculations, finally the angle of each servo motor is calculated. Thus the rotation angle to be achieved for each motor is sent every fifteen milliseconds.

4 Centaur Robot

4.1 Overview

We developed a centaur robot based on the humanoid robot described in the previous section. We prepared two humanoid robots and used one as a front body. For another robot, we only used its lower body as a back of the centaur robot. Then we connected these two parts by a flat plastic board that functions as the shoulder part. Figure 4 shows the centaur robot's appearance.

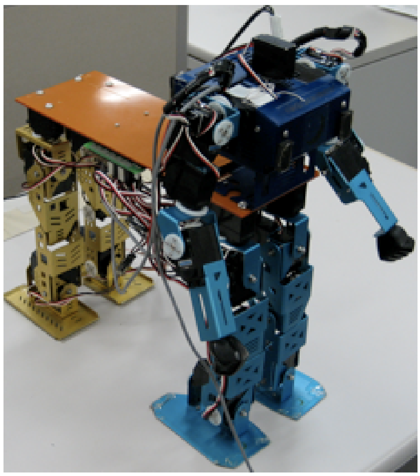


Fig. 4. Centaur robot

4.2 Hardware Construction

Here we explain the robot's hardware construction, as illustrated in Figure 5. Apparently for the front the hardware of the original humanoid robot was used, and for the back only the lower body was used. But a comparison of Figs. 3 and 5 shows that this robot's control structure is somewhat different from the original. Two controllers were used for complete control of the robot. One controls the servo motors required for upper body motions. The other controls the servo motors corresponding to the lower body. Since all the sensors are provided for the upper body, the controller corresponding to the upper body manages all sensor feedback. We adopted these two boards for several reasons. One, by using two boards, one of which controls the motions of the upper body and the lower body, it is possible to separately control the behaviors of the upper body as well as the lower body. For the power supply and battery, both controllers are connected to one battery. Also commands from PC are sent to both controllers.

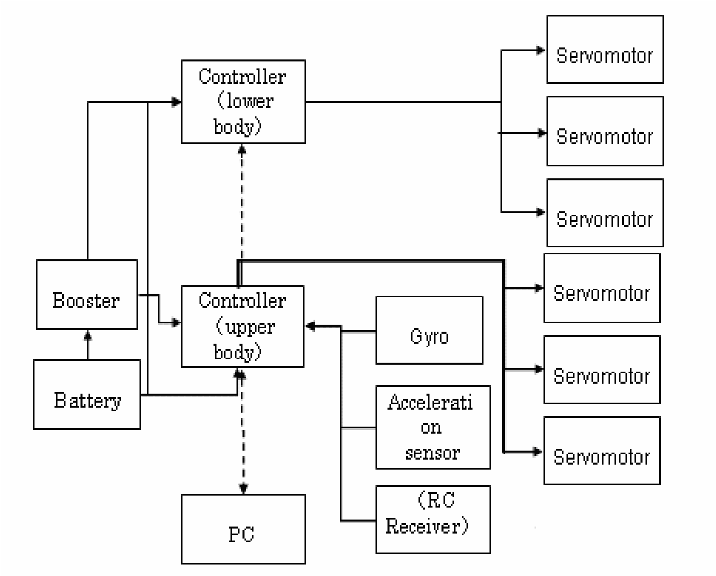


Fig. 5. Hardware construction of centaur robot

4.3 Software Construction

Next, we explain the robot's software construction, as illustrated in Figure 6. The software of the original humanoid controls both the upper and lower bodies together. For the centaur robot, we checked all the original robot's software and separated the software codes into two groups: one that controls the upper body and another that controls the lower body. Thus we reconstructed the whole software. For the upper body, it is unnecessary to carry out calculations for walking. When commands other than a walking command are sent from the PC, it retrieves motion data stored in the memory and sends the necessary rotation angle data to each servo motor. On the other hand software corresponding to the lower body must treat two types of commands as in the case of the original humanoid robot: a command for walking and other commands for additional motions. Also we adopted a method of inserting an arbitrary phase shift between the servo motor control of the front and back legs so that the robot can adopt the most adequate walking motions depending on the walking speed.

By adopting such basic software structure, robot control has the following merits:

(1) The upper and lower body motions can be controlled separately. So far all the motion data developed for achieving various types of humanoid robot motions must be developed to describe the whole body movement. Since the motions of the upper and lower bodies have been separated, we can separately develop two types of motions, and by combining these two types of data, we can generate various kinds of whole body movements for the robot. This idea can easily be applied to the original humanoid robot.

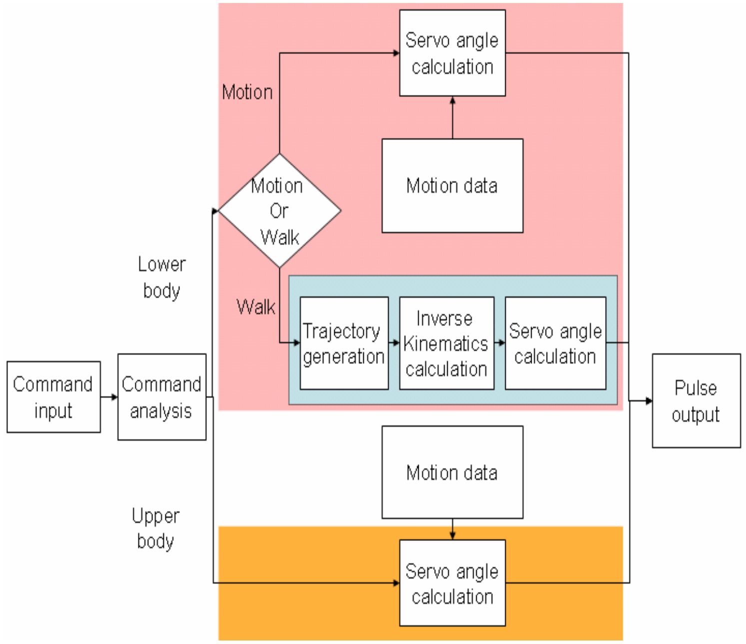


Fig. 6. Software construction of centaur robot

(2) The front and back body movements can be separately controlled. Although it seems natural to let the front lower body and back lower bodies perform identical motions, sometimes it is better to control the two bodies by different body motions. Especially in the case of walking and running motions there would be some differences between these two bodies. For example, for trot type walking there should be a 180^{0} phase shift between the front and the back legs. In the case of gallop running, the front legs and the back legs should move synchronously.

4.4 Evaluation of the Robot

We carried out several experiments to evaluate the motion capability of our centaur robot.

(1) Walking capability

We inserted a phase shift of 0 degree, 90 degree, and 180 degree between the walking motion cycle of the front and back legs. These walking styles correspond to those of animals such as “pace,” “gallop,” and “trot.” We confirmed that the robot could move smoothly with almost the same speed by adopting each of these walking styles. Therefore for walking stability and speed all of the three walking styles perform the same capability.

As a next step, as we expect that one of the applications of this robot would to carry light load, we evaluated the walking stability from a point of carrying a load.

For this we measured the tilt angle of the shoulder when it walks by fixing a gyro sensor on its shoulder and obtaining tilt data from it. We observed the time sequence data of the tilt angle five times for each walking style and averaged the data. Figures 7, 8, and 9 show the obtained data for a phase shift of 0 degree (pace), 90 degree (gallop), and 180 degree (trot).

Figures 7, 8, and 9 show that the tilt angle for pace walking style is larger than other two styles especially at a first step. Thus this waking style is not adequate for carrying a load such as a teacup or a newspaper. When comparing gallop and trot waling styles, although for the first step the tilt angles are almost the same, for the following steps trot walking style shows far better result. In this case of trot walking the front and back legs move in opposite modes. For example, when the front left leg moves forward, so does the back right leg. These results show that the tilt angle

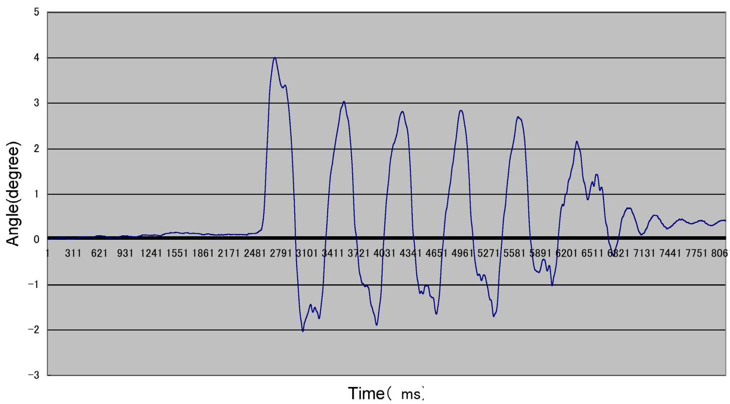


Fig. 7. Tilt angle for “pace” walking style

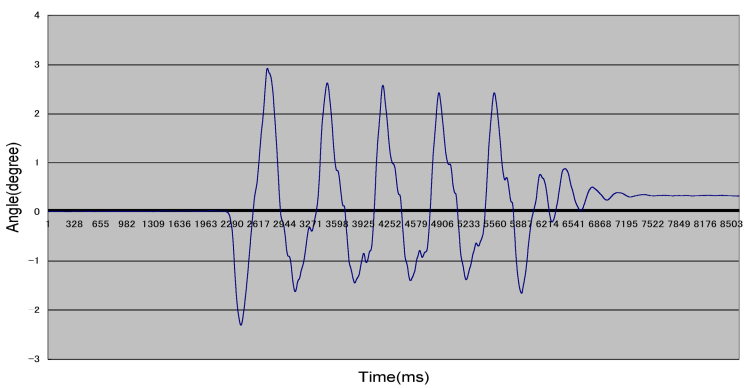


Fig. 8. Tilt angle for “gallop” walking style

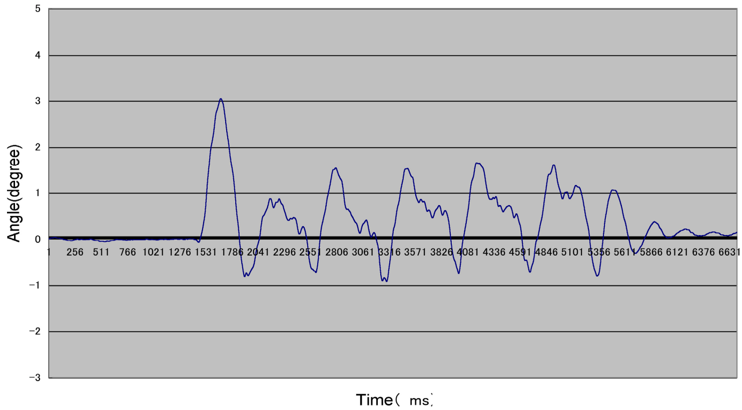


Fig. 9. Tilt angle for "trot" walking style

for trot walking style is lower than the other two and thus more stable when carrying a load. At the same time even in the case of trot, the first step causes a little bit large tilt angle. And this is the issue we have to improve in the further study.

(2) Capability for other motions

We developed various types of human-like motions for the original humanoid robot [8]. An interesting question is which of these motions could work well on the centaur robot. We tried to transfer the humanoid robot motions to this robot and found that most of the motions worked fairly well on this robot. On the other hand, motions including such postures as bending and twisting did not work well or needed modifications. One interesting future research theme is automatically transferring the humanoid robot motions to the motions of four-legged robot such as this robot.

5 Conclusion

In this paper we described the evaluation of walking motion stability of a robot we have proposed and developed. In the previous paper we proposed a new type of robot called "Centaur Robot" that is an integration of two types of robots: humanoid and four-legged. We adopted a humanoid robot with two legs and walking capability as a platform for this new robot. By integrating two of the humanoid robots we easily and successfully developed a centaur robot. In the previous paper we indicated that this robot has a function of stable walking. This time we carried out quantitative evaluation experiments to measure the stability of its waking motion. We confirmed that by inserting a phase shift of 0 degree, 90 degree and 180 degree between the front and back leg motions the robot can stably achieve pace, gallop, and trot walking motions. Then we evaluated these three walking styles from a point of tilt angle of its shoulder and found that the trot walking style is more stable than other styles. Since this robot has merits of both humanoid and four-legged robots, we are also going to evaluate its new capabilities that neither of the two type robots could achieve by themselves.

References

1. http://www.honda/co.jp/ASIMO/
2. Friedmann, M., Kiener, J., Petters, S., Thomas, D., Von Stryk, O., Sakakmoto, H.: Versatile, high-quality motions and behavior control of humanoid soccer robots. In: Workshop on Humanoid Soccer Robots of the 2006 IEEE International Conference on Humanoid Robots, pp. 9–16 (2006)
3. http://www.jp.aibo.com/
4. Golubovic, D., Li, B., Hu, H.: A Hybrid Software Platform for Sony AIBO Robots. In: Polani, D., Browning, B., Bonarini, A., Yoshida, K. (eds.) RoboCup 2003. LNCS (LNAI), vol. 3020, pp. 478–486. Springer, Heidelberg (2004)
5. Kerepesi, A., Kubinyi, E., Jonsson, G.K., Magnusson, M.S., Kiklosi, A.: Behavioural Comparison of Human-Animal (Dog) and Human-Robot (AIBO) Interactions. Behavioural Processes 73(1), 92–99 (2006)
6. Ishiguro, H., Ono, T., Imai, M., Kanda, T.: Development of an interactive humanoid robot "Robovie" -An interdisciplinary approach. In: Jarvis, R.A., Zelinsky, A. (eds.) Robotics Research, pp. 179–191. Springer, Heidelberg (2003)
7. Tsuda, S., Oda, Y., Shinozaki, K., Nakatsu, R.: Concept and Architecture of a Centaur Robot. In: Ma, L., Rauterberg, M., Nakatsu, R. (eds.) ICEC 2007. LNCS, vol. 4740, pp. 381–388. Springer, Heidelberg (2007)
8. 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)
9. Shinozaki, K., Oda, Y., Tsuda, S., Nakatsu, R., Iwatani, A.: Study of Dance Entertainment Using Robot. 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)
10. Shinozaki, K., Iwatani, A., Nakatsu, R.: Concept and Construction of a Robot Dance System. The International Journal of Virtual Reality 6(3), 29–34 (2007)

Analysis of Japanese Folktales for the Purpose of Story Generation

Takaaki Kato[1], Takenori Wama[1], Koji Miyazaki[1], and Ryohei Nakatsu[2]

[1] Kwansei Gakuin University, School of Science and Technology
2-1 Gakuen, Sanda, Hyogo, 669-1337 Japan
{bqz61917,scbc0057,miyazaki}@kwansei.ac.jp
[2] National University of Singapore, Interactive & Digital Media Institute
Blk E3A, #02-04, 7 Engineering Drive 1, Singapore 117574
idmdir@nus.edu.sg

Abstract. We are developing an interactive folktale system in which users can enjoy both the generation of Japanese folktales as well as interaction with the system. In order to generate any kind of Japanese folktale, the system must give each character appearing in the folktales the ability to achieve a sufficient number of motions. To determine the number of motions that need to be prepared for the system, we carried out an analysis of the verbs that appear in Japanese folktales. Based on the analysis of 50 representative Japanese folktales, we found that the total number of verbs used is more than 2,000. By deleting the overlap of verbs in different folktales, while giving consideration to their meanings in actual usage, the number decreased to about 900. In addition, by restricting the verbs to those that can be expressed by computer graphics, the total number of verbs further decreased to about 100. Then we carried out experiments on generating various scenes of Japanese folktales by using several motion sets, each of which is a subset of the 100 motions corresponding to the reduced verbs. Finally, we evaluated the least number of animations needed to generate animations of sufficient quality.

Keywords: folktale, story generation, verb, animation.

1 Introduction

The authors are developing an interactive story generation system [1][2]. We adopted Japanese folktales as a sample genre for story generation. Each culture has its own folktales, and children grow up listening to their folktales or watching animations of folktales on TV or in movies. Consequently, folktales have been playing an important role in the education of children in every culture. Naturally, this holds true for the Japanese folktales recounted to the children of Japan. Therefore, we can expect huge educational potential from an interactive system that, going beyond watching or listening to folktales, gives joy to children through interaction with the folktales' characters. When we consider children as the users of such an interactive system, the story should be enchanting enough to keep the interest of children. At the same time, the quality of computer graphics is another key issue for such a system to be accepted by

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 67–76, 2008.

people. In particular, the types of motions characters could achieve in such an interactive system have a strong impact on the impressions held by children. Therefore, an evaluation of such an interactive system, from the viewpoint of how the system can entertain children, should take these factors into consideration.

When the system generates interactive stories with computer graphics of good quality, it is necessary for the system to prepare an adequate amount of motion data in advance to generate any kind of action needed by the characters in a story. Since verbs are used in stories to reflect action, it is worthwhile to analyze the verbs used in Japanese folktales to determine what kinds of verbs and how many verbs appear in them.

Based on the above considerations, we analyzed 50 representative Japanese folktales. In this paper, we describe the details of our analysis. The results would be helpful in determining the maximum number of character motions that need to be prepared for the motion database of an interactive folktale system. Another interesting issue would be clarifying the least number of character motions that need to be prepared for the generation of Japanese folktale animations having adequate quality. To determine this, we carried out an evaluation experiment. Based on the analysis of verbs in Japanese folktales, we have prepared several motion sets that each have a different number of motions for the characters in the folktales. Then we asked subjects to watch multiple animations, generated using each of these motion sets, and to evaluate each animation. We also describe the details of the experiments.

2 Related Works

An interactive system, and especially one with a story, has a strong power to attract people [3]. This has been proved by the fact that video games have been so successful. Furthermore, it is expected that this power of interactive stories could be applied to education or e-learning. Based on this expectation, we are developing the Interactive Folktale System [1][2]. Several scenes generated by the system are shown in Fig. 1.

One of the most important issues for interactive stories is controlling their development, which has been actively studied as "Interactive Storytelling" [3][4][5][6]. The key issues in interactive storytelling include the generation of autonomous

Fig. 1. Examples of scenes in Interactive Folktale System

interactive characters and the emergence of storylines based on interactions between virtual characters and users, as well as those among the virtual characters. This is undoubtedly the final target of interactive storytelling. The problem with present interactive storytelling is that, since the generation of sophisticated autonomous characters is so complicated, it is difficult to maintain the consistency of the generated story for more than several minutes [7]. On the other hand, in most of these works all of the necessary motions of the characters in the system have been prepared beforehand and stored in the system, and little attention has been paid to the types and numbers of motions characters could achieve. As the number of stories and the length of the stories increase, this becomes an important issue. Therefore, it is important to study what types of motions and how many motions need to be prepared beforehand. However, so far few studies have investigated this topic.

For story generation, it is important to study the basic structure of stories. One of the most famous studies in this area is that by Vladimir Propp [8]. He analyzed more that 100 Russian magical folktales and revealed several interesting facts. One is that all the story lines consist of concatenations of sort stories. Our Interactive Folktale System is constructed based on this basic concept [1][2]. Once we could prepare motion database that contains enough number of motion animation sets, the system can generate any kinds of Japanese folktales.

3 Methodology of Analysis

Various books have been published in Japan on Japanese Folktales. Among them, we selected a book as a reference source that contains 100 representative Japanese folktales [9]. A variety of folktale animations have been developed and broadcasted on TV in Japan based on this book, so we considered it a valid basis for our analysis. Fifty representative folktales were chosen from these one hundred folktales. In our analysis method, we extracted all of the verbs appearing in these 50 folktales and then counted the instances of each while considering various factors of their use. In the following, we describe the procedure followed in this analysis.

3.1 Step 1 Verb Extraction

As a first step, all of the verbs that appear in these 50 folktales are extracted. If the same verbs appear in one story, we merge these verbs and count them as only one instance.

3.2 Step 2 Deletion of Overlaps

In the next step, if the same verbs appear in different stories, we also merge them and count them as only one verb. For example, the same verb "say" appears in different Japanese folktales such as "Urashima-Taro (Fisherman Urashima)" and "Momo-Taro (Peach Boy)," so we merge these usages.

3.3 Step 3 Deletion of Overlaps Taking *meaning* into Consideration

We then carry out the merging of verbs with the same meaning. For example, such verbs as 'speak' and 'talk' or 'go' and 'come' would be expressed by the same kind of CG animations. Therefore, these verbs can be merged.

3.4 Step 4 Classification of Verbs into Several Categories

As the next step, we consider more deeply the meaning of each verb. Here, we try to classify all of the extracted verbs into several categories from the viewpoint of meaning. For this, we prepared six categories: (1) verbs corresponding to action, (2) verbs corresponding to situation, (3) verbs corresponding to action depending on objects, (4) verbs corresponding to emotion, (5) verbs describing natural phenomena, and (6) other verbs. For example, the verb 'go' is classified into the first category. Such verbs as 'think' and 'decide,' corresponding to mental behaviors, are classified into the second category. Such verbs as 'hold' or 'hit' are classified into the third category. Since actions based on computer graphics would vary depending on the object of such verbs, we prepared this category. For example, in the case of 'hold,' computer-graphics animations would vary depending on whether this means holding a light object or a heavy one, a large object or a small one, etc. Verbs describing emotional states such as 'get angry' or 'joyful' are classified into the fourth category. Such verbs as 'rain' or 'snow' that are used to describe natural events or phenomena are classified into the fifth category. The last category was prepared to include those verbs that cannot be classified into the above categories. For example, such verbs as 'stop doing something' or 'do not do something' are classified into this category.

3.5 Step 5 Detailed Analysis and Reduction of Verbs Considering Computer Graphics Expression

Based on the results of step 4, we carry out more detailed analysis of the verbs. Since the first category is most important when preparing CG-based animation, we focus on the verbs in this category. This type of verb can be classified into two subcategories: verbs expressing the movement of a character and verbs expressing actions it takes toward objects or other characters.

The former contains such verbs as 'return,' 'accompany,' 'come back,' and so on. These verbs can be expressed by computer graphics in such actions as walk, run, swim, and fly. Since all of these verbs imply actions of movement from one place to another, it is not necessary to prepare special animations such as 'return' or 'come.' Instead, we need only prepare the animations corresponding to each mode of movement, such as walk, run, and so on. Now, let us consider the latter cases. Here, it is necessary to prepare a special animation for each of the actions such as cook and sew. However, if a character does action such as moving around in a kitchen, we can easily imagine it is cooking. Furthermore, in the case of sewing, if a character already has a needle and a cloth, a simple action such as moving the hands in some way would be close enough.

At the same time, what the CG animated characters need to do depends on the particular model of a character. If the model were simple, the animations the character would be able to do would be significantly restricted. However, if we reduced the number of animations too much, then the actions of the characters would become unnatural. This means that the kinds of animations we have to prepare for CG characters depend on the characteristics of the characters as well as the kind of stories we have to regenerate. Consequently, we have to consider what would be the least acceptable number and fewest kinds of character actions in the case of Japanese folktales.

Taking these issues into consideration, we carefully checked the verbs in the first category (verbs corresponding to actions) and reduced the number of verbs by merging them. We then prepared computer graphics animations corresponding to each of the merged verbs.

4 Analysis Results

Based on the methodology described in the previous chapter, we analyzed the verbs that appear in the 50 representative Japanese folktales [9]. Here we describe the analysis results. Figure 2 illustrates the result of step 1 (results of original verb data), step 2 (result after simple overlap deletion), and step 3 (result of deletion of overlap taking meaning into consideration). The horizontal axis corresponds to the number of folktales and the vertical axis corresponds to the accumulated number of verbs. The accumulated number of verbs in each step is 2324, 1613, and 884. Figures 3, 4, 5, and 6 illustrate the accumulated number of verbs in each of four verb categories, i.e., action verb, object-dependent verb, situation-based verb, and emotion-related verb. The total number of verbs corresponding to each of these categories is 306, 137, 202, and 64, respectively.

In Fig. 2, the trend before deleting overlap shows a simple increase. On the other hand, after step 2, the increase ratio became smaller, and after step 3, a saturation tendency became visible. Figures 3–6 show how the number of verbs in each category increases depending on the number of folktales. Although we have to further investigate the trends in these verbs by increasing the number of analyzed folktales, each graph already shows a gradual saturation.

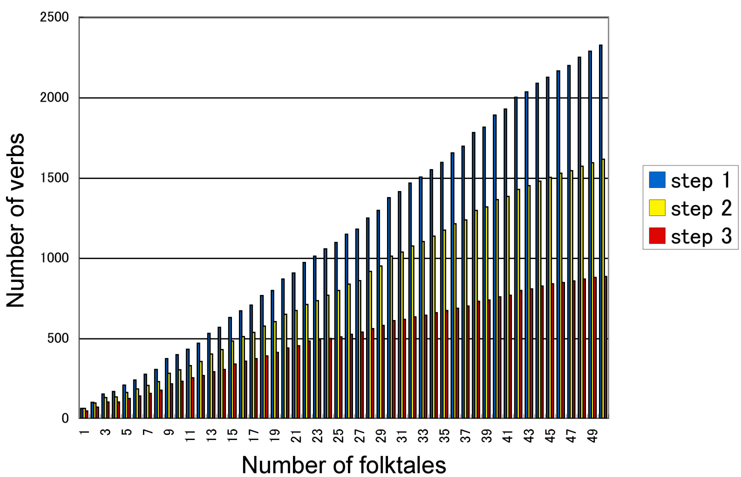


Fig. 2. Number of folktales vs. number of verbs

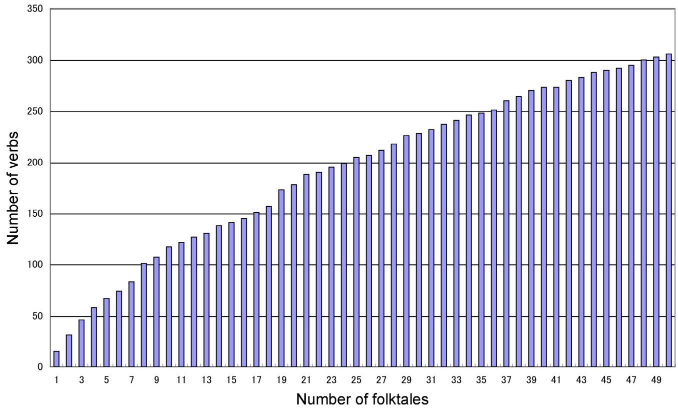


Fig. 3. Number of folktales vs. number of verbs (action verbs)

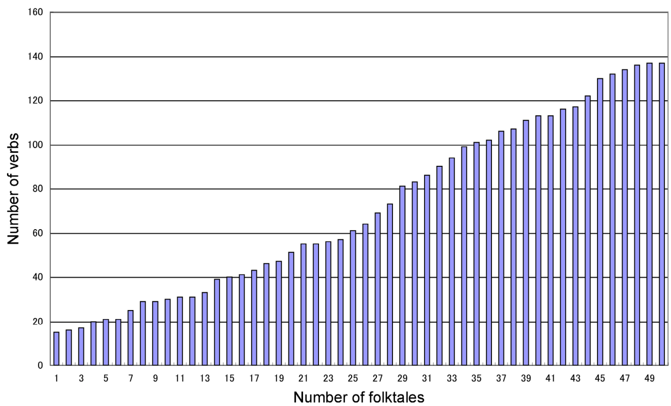


Fig. 4. Number of folktales vs. number of verbs (object-dependent verbs)

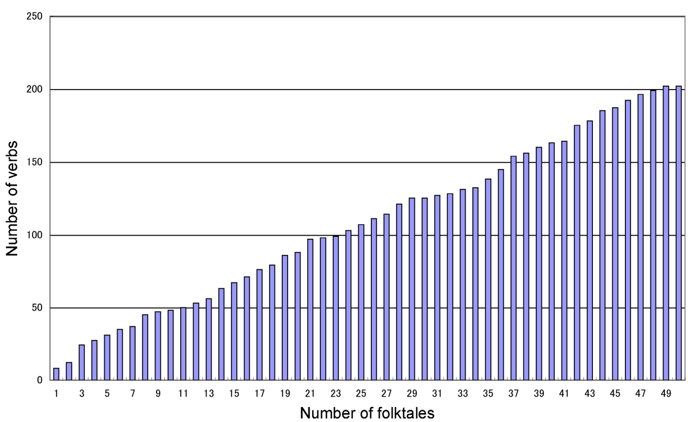


Fig. 5. Number of folktales vs. number of verbs (situation-based verbs)

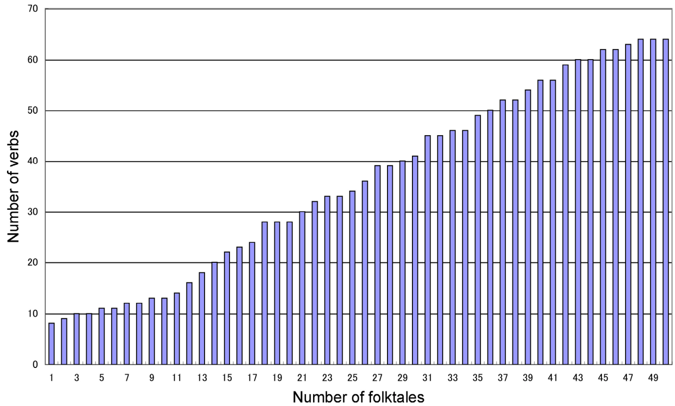


Fig. 6. Number of folktales vs. number of verbs (emotion-related verbs)

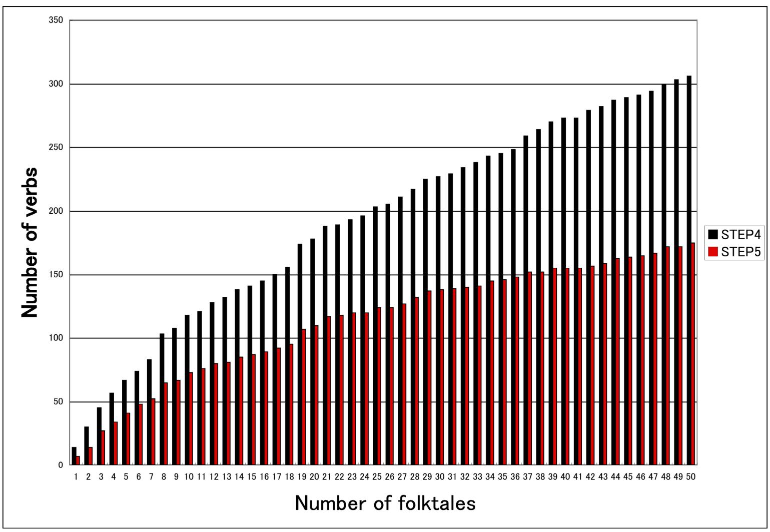


Fig. 7. Number of folktales vs. number of verbs (Step 5)

Figure 7 illustrates the results of step 5. The number of verbs after step 5 is reduced to 175. Moreover, the effect of saturation became clearer. This graph shows that about 170–200 actions should be prepared as computer graphics animations in order to generate a sufficient number of Japanese folktales.

5 Evaluation Experiment

Based on the results of the above analysis, we carried out evaluation experiments to determine how many character actions would be sufficient to generate animations of Japanese folktales with high enough quality. We selected two representative stories from the 50 stories (story 1 and story 2) to generate as computer graphics animations.

Among 170 total verbs, only about 50 verbs appear in these two stories. Among these 50 verbs, we carefully extracted ten, twenty, and forty representative verbs and called them group 1, group 2, and group 3, respectively. Story 1 consists of eleven scenes, where seven of them are generated using the three types of motion sets described above. Story 2 consists of thirteen scenes, ten of which are generated using the three motion groups. Figure 8 shows a comparison of different motions in the same scene. In Fig. 8 (a), the scene shows the character of an old man running. In the case of Fig. 8 (b), however, the motion "run" is not included in the motion group, so the motion "walk" is used instead of run.

Fig. 8. Comparison of different motions in the same scene

5.1 Evaluation Experiment 1

In this experiment, we evaluated each scene generated using the three motion sets from the viewpoints of naturalness and fun. Ten subjects, aged around twenty years, were selected and asked to watch three types of generated scenes shown in random order. Then they were asked to evaluate these animations from the viewpoints of "naturalness" and "fun." The evaluation scores consist of five rankings; 1: worst, 2: worse, 3: neutral, 4: better, and 5: best. The results of the evaluation test are shown in Table 1.

Table 1. Results of evaluation experiment 1

	Group 1 (10 motions)		Group 2 (20 motions)		Group 3 (40 motions)	
Evaluation item	Naturalness	Fun	Naturalness	Fun	Naturalness	Fun
Mean value	2.40	2.61	2.71	2.85	2.81	2.93
Variance	0.16	0.10	0.28	0.32	0.12	0.18

These results show that as the number of actions prepared as character animations increases, the evaluation scores for both naturalness and fun also increase slightly. However, the differences are not large in light of the variances. Therefore, these results show that for naturalness and fun the number of actions is not so important. On the other hand, it can be easily conjectured that for the quality of the generated animations, the story itself would have a big effect. Since our evaluation involved only short scenes, the effect of story may not have been fully expressed.

5.2 Evaluation Experiment 2

To clarify the effect of story, we carried out a second experiment. In this case, the subjects are asked to watch the scenes generated sequentially according to the story. All of the other experimental conditions are the same. The results of the experiments are shown in Table 2.

Table 2. Results of evaluation experiment 2

	Group 1 (10 motions)		Group 2 (20 motions)		Group 3 (40 motions)	
Evaluation item	Naturalness	Fun	Naturalness	Fun	Naturalness	Fun
Mean value	2.13	2.53	2.93	3.07	3.6	3.73
Variance	0.52	0.25	0.6	0.2	0.91	0.2

These results show the following. In the case of animations using group 1 motion sets, evaluation scores for both naturalness and fun are around two and thus not adequate for generating animations. For the animations using group 2, both naturalness and fun achieved scores around three. For the generation of folktales, this implies that at least twenty basic motions are necessary for maintaining the minimum acceptable level of quality. On the other hand, the animations using the group 3 motion set earn a relatively high score between three and four. This suggests that it would be possible to generate relatively good quality by using forty motions. At the same time, however, it is still necessary to carry out further experiments using larger numbers of motions to determine where quality becomes saturated.

6 Considerations and Future Work

Here, based on the results of the analysis and evaluation, we discuss additional considerations as well as several points that should be further studied.

As described in the previous section, the animations achieved by CG characters would depend on the characteristics of the characters themselves. If characters are based on a precise model, such as having five fingers on each hand and the capability to mimic human actions, then we need to prepare all of the actions corresponding to all of the verbs. On the other hand, however, it the models of the characters are simple enough that they don't have all ten fingers, then the actions using fingers should be expressed in some other way.

Other than the verbs corresponding to actions, let's consider verbs corresponding to emotions. Such emotional verbs as 'cry' or 'get angry' can be expressed either by facial expressions or body motions or combinations of these. In any case, the facial expression plays the most important role. Therefore, verbs corresponding to emotions are somewhat special.

In the case of situation-based verbs corresponding to mental behaviors such as think, decide and so on, it is fairly difficult to express these verbs only by body motions. Therefore, it is necessary to introduce narrations in addition to body motions. Moreover, the motions corresponding to these actions largely depend on culture. Therefore, the preparation of animations for these verbs would require extraordinary care.

As future work, we first have to increase the number of folktales for analysis to further clarify the trend of each type of verb. In addition, we intend to carry out further evaluation experiments by using more folktales.

6 Conclusion

We have analyzed the verbs appearing in 50 representative Japanese folktales for the purpose of developing an automatic story-generation system. Although we need to carry out further research, so far we have found that there is some limit on the number of animations that need to be prepared for our system. This limitation will closely correspond to the particular actions of the characters in the folktales.

References

1. Miyazaki, K., Nagai, Y., Bosser, A.G., Nakatsu, R.: Architecture of an Authoring System to Support the Creation of Interactive Contents. In: Harper, R., Rauterberg, M., Combetto, M. (eds.) ICEC 2006. LNCS, vol. 4161, pp. 165–174. Springer, Heidelberg (2006)
2. Miyazaki, K., Nagai, Y., Wama, T., Nakatsu, R.: Concept and Construction of An Interactive Folktale System. In: Ma, L., Rauterberg, M., Nakatsu, R. (eds.) ICEC 2007. LNCS, vol. 4740, pp. 162–170. Springer, Heidelberg (2007)
3. Murray, J.: Hamlet on the Holodeck. The Free Press (1997)
4. Swartout, W., et al.: Toward the Holodeck: Integrating Graphics, Sound, Character and Story. In: Proceedings of the Autonomous Agents 2001 Conference (2001)
5. Mateas, M., Stern, A.: Socially Intelligent Agents: The Human in the Loop. In: AAAI Fall Symposium (2000)
6. Young, R.M.: Creating Interactive Narrative Structures: The Potential for AI Approaches. In: AAAI Spring Symposium in Artificial Intelligence and Interactive Entertainment. AAAI Press, Menlo Park (2000)
7. Charles, F., Cavazza, M.: Exploring Scalability of Character-based Storytelling. In: Proceedings of ACM AAMAS 2004 (2004)
8. Prop, V.: Morphology of the Folktale. University of Texas Press (1968)
9. 100 Japanese Folktales, Kodansha (in Japanese) (1997)

Strategic Path Planning on the Basis of Risk vs. Time

Ashish C. Singh and Lawrence Holder

School of Electrical Engineering and Computer Science
Washington State University
Pullman, WA 99164
ashish.singh@ignitionflorida.com, holder@wsu.edu

Abstract. The selection of path in an urban combat setting determines the survival to a greater extent. In this paper we propose an algorithm that finds strategic paths inside a map with a set of enemies without using predetermined waypoints. The strategic path calculation is based upon the hit probability calculated for each enemy's weapons and the risk vs. time preference and it is done at multiple levels of abstractions to address trade-off of efficiency and accuracy and the strategic path calculation minimizes both time and risk as per mission objectives.

Keywords: Strategic Path Planning, Visibility Algorithm, Risk, Time, Non-player character (NPC), Line-of-Sight (LOS), Heuristic Space Search (HSS), Military perations on Urban Terrain (MOUT).

1 Introduction

In this work we present a technique by which without using the fixed set of waypoints we can compute almost unlimited variation of paths based upon the path's risk evaluation. Thus, from the game designer's standpoint it can add to unlimited variations to the gameplay without requiring any manual marking of *navigation* or *cover point* on the map.

We have developed a strategic path planning algorithm that is based upon in-depth risk evaluations along all the possible paths that can lead to the goal. We use the hit probability for calculating the risk involved on a path. The risk calculation takes into account all the enemies.

In our work Risk is defined as the ability to shoot the player in terms of hit probability (HP). Each weapon has a different hit accuracy, rate of fire and hit ratio per bullet fired. We used weapon details [1] to obtain a HP based on distances from a set of enemies. Each weapon has a different HP. The enemy's ability to shoot the agent depends upon three factors: (1) the agent's visibility from the enemy's location, (2) the distance from the enemy, and (3) the lethality of the enemy's weapon.

A strategic path is a trade-off between the time of traversal and the risk along the path. Not all the areas along all the possible paths are completely covered. Therefore the risk evaluation must consider all of the three components of risk. We developed the strategic path computational model using these techniques in the context of a *MOUT* scenario within the Quake3 first-person shooter game.

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 77–87, 2008.

In section 2, we compare our work with the research work done in this field. Section 3 discusses our testbed environment, the agent's interface and graph conversion of the map. Section 4 contains important concepts of strategic path computation. Section 5 presents our experimental results, and section 6 presents our conclusions.

2 Related Work

Path planning and collision prevention for single and multiple players has been extensively studied [2]. But strategic path planning has not received as much study. Shortest path planning can be done on waypoints by applying the A* algorithm [5]. But this approach neglects the strategic importance of waypoints.

Using the *BitStrings* technique Liden [4] has done strategic path planning to exhibit MOUT tactics like *flanking*. In Liden's work risk has not been studied in detail. Risk is defined by the ability of an enemy to kill the agent. Liden makes three simplifying assumptions. First, a distant enemy is considered equally risky compared to a short distant enemy and also the variation in firepower is neglected. Third, *BitStrings* can only be used for a fixed set of waypoints. In real 3D environments, the visibility complexity increases and a set of fixed waypoints cannot accurately address the strategic importance of visibility and also computed paths are limited in count. Our work addresses each of these assumptions to yield more strategically-realistic paths.

3 Urban Combat Testbed

Our experiments were performed using the *Urban Combat Testbed (UCT)* [9]. UCT is a mod of the *Quake3* first-person shooter game. The agent program exchanges percepts and actions with the UCT using a shared memory interface that allows lower communication latency and lower computational burden on the game engine.

The percepts are of two types: dynamic and static. The dynamic percepts include information about current location, health, weapons and ammunition, i.e., these percepts are meant to change with the game play. There are 33 different dynamic percepts related to the player, 11 different percepts about entities which include opponents if they are present, 4 different percepts about weapons, and all the different dynamic objects. The static percepts contain the map information. The static map information is passed to the agent using an XML description from UCT's Static Spatial Perception Service (SSPS).

The agent program can choose from 29 different actions that can be sent to the game. The actions are of very primitive form (WALK_FORWARD, TURN_RIGHT, TURN_LEFT, etc). The agent program can write actions into the assigned shared memory. The UCT reads these actions and executes them accordingly. For our study, we used the *Reykjavik map* (figure 1- left), which models an urban area.

3.1 Areas and Gateways

The walk-able surfaces in the map have been defined as *areas* [9]. These areas are 3D convex polygons. Areas have been constructed from the 3D *brushes* defined in a Quake3 map. Figure 1 (Right) shows the areas computed for the map in figure 1 (Left).

All walk-able areas are connected using *gateways*. Gateways also contain information about the type of action required to cross the gateway from one area to another area (actions like Jump, Walk, Fall, etc.). World coordinates of *areas*, *objects* and *gateways* are initially parsed using an XML file. From the dynamic and the static percepts the agent calculates the current *area* information. For traversing into another area the agent finds the *gateway* information corresponding to the present *area* and the desired next *area*. The agent sends the relevant actions in order to cross the found *gateway* to the next *area*.

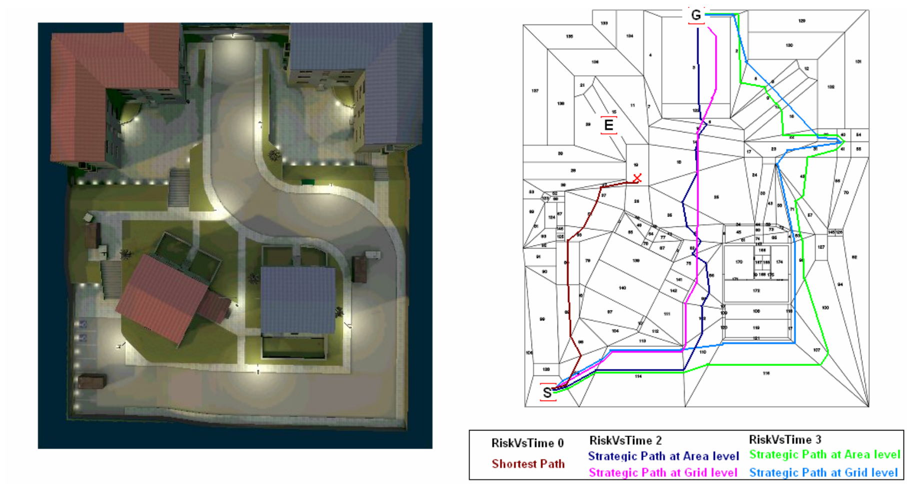


Fig. 1. (Left) The Reykjavik map [3,7] in UCT and (Right) Avg. distance of shortest paths and strategic paths over Area and Grid level

3.2 Area Connectivity Graph

Finding a path between areas becomes a problem of finding a path in the graph constructed from the connectivity information of the map. The connectivity between areas resembles edges between vertices. As in figure 2, areas and their connectivity can

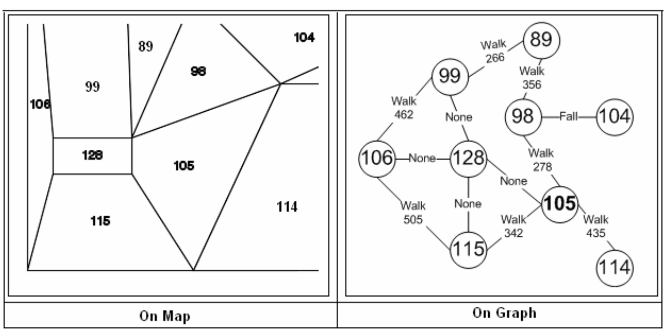


Fig. 2. Graph representing a set of Areas and Gateways

be formalized as vertices and edges of a graph. The Euclidean distances between areas become weights on the edges. In our strategic path calculation we modify these weights to incorporate risk, and then use the same shortest path algorithm to find a strategic path.

3.3 Visibility and 3D Volume Search

Polygonal areas are used both for visibility calculations and also for walk-able path calculations. We have developed the Heuristic Space Search (HSS) technique [6] that can limit the number of visibility tests to a small number of 3D areas and objects. In this technique we index the complete 3D map to a small 3D pointer array, where each pointer points to a list of 3D volumes occupying a fixed 3D space.

In strategic path calculations *brushes* that make an object have been grouped to represent one object, and similarly a set of *brushes* meant to represent a walk-able area represent one area (concept of area abstraction). Therefore, the abstracted areas and objects are larger 3D volumes, which reduce 3D related computations. The HSS constructed from these abstracted 3D volumes is better indexed. Thus, the HSS minimizes the potential 3D volumes for visibility tests and other geometric search tasks (e.g., point containment). Theoretically on perfect indexing (by using a very small HSS edge length) the visibility and other related calculations will be constant time operations.

When compared to the *Binary Space Partitioning (BSP)* technique, the *Heuristic Search Space* technique will directly reach the potential candidates while the BSP technique will search through the root node to the potential set of 3D volumes (*brushes)* by comparing *log(n)* partition planes where *n* is the total number of partition planes.

4 Strategic Path Computation

The strategic path planning is done at two levels. At the *Area level*, a higher level of abstraction, the computation gives the *areas* to walk over. At the *Grid level*, a higher level of detail, the computation gives the within-area *grid points* to walk through after reaching a selected area.

For the strategic path computation movement across the risky areas in the map have been penalized by computing a *Meta-Weight* that is based upon the *hit probability* (from all the enemies) and the given *RiskVsTime* factor. Thus priorities based upon enemies' lethality and preference for safety with respect to time of traversal can be considered together for strategic path planning. The strategic path computations have been done by modifying the weights in the weight matrix *W* of the connectivity graph of *areas* and then within each area over a set of *grid points*. We use *Dijkstra's* algorithm (computational cost: $O(|E|\log|V|)$ if a binary heap is used and $O(|E| + |V|\log|V|)$ if a Fibonacci heap is used) to compute the shortest path, which is the *strategic path*. Given the use of the HSS technique that theoretically allows visibility and area search related operations in constant time, the computational complexity of the strategic path computation is bounded by the computation complexity of the shortest path computation.

4.1 Forming a Small Set of Areas

All the walk-able surfaces in the *Reykjavik* map (figure 1 - Left) have been manually converted into a small set of large convex polygonal areas (figure 2). The *Reykjavik* map contains 125 open areas and 50 closed areas (inside buildings, etc.). These areas are connected using *gateways* which contains connection information and also the action required to move from one area to another (walk, jump, fall, none, etc.). Thus, for a small set of areas the graph representation and further application of the shortest path algorithm are computationally feasible. The abstraction of the map into a small set of areas adds to simplicity and efficiency for strategic path computation at the area level.

4.2 Hit Probability Calculation

From a start area to a goal area there can be many paths. A path is defined by a set of walk-able polygonal areas. These areas can be visible to enemies. The risk factor of a path is determined by calculating the *total hit probability* for the entire distance covered on that path.

The *hit probabilities* have been calculated from the realistic data obtained from [1]. The realistic data gives a rough conversion of distances to static hit probability for various weapons computed for a standing soldier. We convert the given static hit probability to a dynamic hit probability by considering it to be 0.25 times the static hit probability. When the agent or the enemy is moving, it will be harder to hit the agent, so taking a fraction approximates the dynamics of the situation. We considered 3 types of weapons: assault rifle (AK 47), sniper rifle (SKS-84M) and sub-machine gun (MP5).

The strategic distance between an enemy and the agent depends on how dangerous the enemy is. For example an enemy with a sniper rifle is considered more dangerous than an enemy with a lower-power weapon. Thus, the strategic path computation takes into consideration the variation in the tactical distances of the enemies.

4.2.1 Checkpoints

As shown in figure 3 (Left), a point on a path where the risk is computed has been defined as a *checkpoint*. In order to calculate the risk associated with a distance, the *checkpoints* are uniformly distributed along the path at a fixed interval. The risk associated with a distance of walk is computed as the total hit probability of receiving one hit along the *checkpoints*. In the equation 1 HP_{total} represents the total *hit probability* over the given path, and HP_i represents the *hit probability* for *checkpoint* P_i. The total *hit probability* is computed as:

$$HP_{total} = HP_1 + HP_2(1-HP_1) + \ldots + HP_n(1-HP_{n-1})(1-HP_{n-2})\ldots(1-HP_1) \tag{1}$$

Here $(1-HP_1)$ is the probability of not getting hit at checkpoint P_1 and $HP_2(1-HP_1)$ is the probability of only being hit at checkpoint P_2. Similarly, $HP_n(1-HP_{n-1})(1-HP_{n-2})\ldots(1-HP_1)$ represents the probability of only taking a hit at checkpoint HP_n (after not taking hits over the previous *checkpoints).* The *total hit probability* HP_{total} represents the probability of receiving one hit at one of the *checkpoints*.

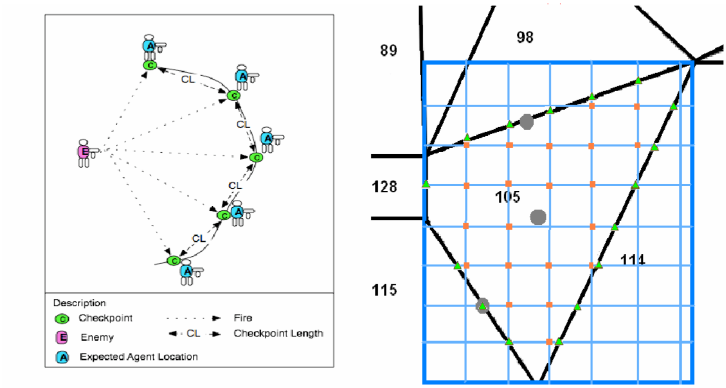


Fig. 3. (Left) Checkpoint distribution over a path and (Right) the strategic path at Grid level for Area-105

For the strategic path computation, initially the risk for each pair of neighboring areas is computed. Here, the risk is the total hit probability along the *checkpoints.* The first *checkpoint* is allocated to the center of the start area and the rest of the *checkpoints* are distributed at fixed intervals (e.g., 1m) along the path to the center of the stop area. This path goes through the connecting gateway of the two neighboring areas. Thus, the risk over a pair of areas is the total hit probability across the distributed checkpoints. The *hit probability (HP)* over this path is used for the Meta-Weight computation.

4.3 Risk vs. Time Preference Factor

Risk is attributed to the probability of a hit. In order to succeed on a mission the agent must maintain a minimum health and minimize health damage. This can be done by taking a route that keeps the agent hidden from most of the threats on the map. But not all the paths are threat free.

As per mission objectives the agent may want to reach a goal location as soon as possible and must take the shortest route towards the goal. But the shortest route may contain threats. Thus, the agent must make a trade-off in selecting a path that can minimize risks and time. Depending upon the mission objectives the preference for the shortest route compared to the preference for safety may vary. Thus, in order to maintain a good balance between the safest path and the shortest path the agent must define its risk vs. time preference factor. A high value will prioritize safety and a low value will prioritize time of traversal.

The *RiskVsTime* factor penalizes an exposed path by linearly increasing the cost of traversal of the path (Euclidean distance). The *RiskVsTime*=1 factor will double the cost of traversal over an area with hit probability 1. The *RiskVsTime*=0 factor will keep the cost of traversal over that area unaffected. As the *RiskVsTime* factor increases the strategic path computation tends toward more safety with a trade-off of longer route to the goal.

4.4 Meta-weight Calculation

After determining the risk vs. time preference and the hit probability on the connecting distance across the neighboring areas, we can compute a meta-weight. This meta-weight is multiplied to each of the weights in the weight matrix. This meta-weight represents a penalizing factor meant to symbolize the extra cost for being exposed to enemies:

$$\textit{Meta-Weight} = (HP * \textit{RiskVsTime} + 1) * \textit{Euclidean Dist.} \tag{2}$$

Euclidean distance is the distance between area centers through the connecting gateway, and *HP* is the total *hit probability* over this Euclidean distance. The equation 2 takes care of the tactical priorities between more dangerous and less dangerous enemies as well as maximizing the strategic distance from the enemies based upon the *RiskVsTime* factor.

4.5 Strategic Path at Grid Level

After the strategic path at the Area level is computed, the strategic path at the Grid level (higher level of detail) is computed for each selected area while traversing the strategic path at Area level. As shown in figure 3 (Right) a selected area is further divided using a grid formation with fixed *grid unit length*. A larger *grid unit length* means more detail and more computational cost. This computation gives the within-area *grid points* to walk through for each selected area. This is done during the path traversal, so it takes into consideration any enemy movement. A selected area is subdivided using a Grid formation, and a strategic path is computed over the *grid points*.

The strategic path at Grid level computation is a two step process. First, a *gateway point* is computed on the gateway between the current area and the next selected area. The *Gateway point* is computed with the same strategic path principle, and in addition, the distance from the goal location is minimized and paths are made smoother. In the second step, *grid points* are distributed on the current area, connected to each other and to the two *gateway points* according to adjacency. The strategic path between the two *gateway points* is then computed as the strategic path at the Grid level. This technique is applied over all the selected areas (selected by the strategic path at Area level) while traversing each of the selected areas.

5 Experimental Results

We present experimental results for both *out-game* and *in-game* trials. *Out-game* trials are meant to simulate *in-game* trials so that we can perform a more systematic analysis of our approach. We performed the *Out-Game* trials using the realistic weapon details [1]. By using the *checkpoint* technique, we computed the hit probability of the generated strategic paths. The *In-Game* trials were done using the Urban Combat Testbed (a modification of Quake3). For each experimental condition (start area, goal area, enemy area), we ran 10 trials and averaged the results. For example in an experiment where the agent was hit 4 times and successfully reached the goal without getting a hit 6 times, the hit probability is 0.4. We performed the in-game trials to validate the accuracy of the out-game computation.

We computed the difference between the *in-game* and *out-game* trials (*hit probabilities* for the traced paths) with *RiskVsTime*=10 and an Assault Rifle for the enemy's weapon. For the strategic path at Area level the difference was 0.17761, and for the strategic path at Grid level the difference was 0.1896. These differences between the *in-game* and the *out-game* trials for the strategic paths are good considering the differences between real-world weapon performance and *Quake3*'s implementation of a similar weapon (similar but not the same).

The *out-game* trials were based on the realistic risk evaluation computed from available weapon details and were free from the implementation details of the Quake3 game engine. We performed *out-game* trials varying the *RiskVsTime* factor from 0 to 30. We observed that the computed strategic paths (Area and Grid levels) for a high *RiskVsTime* factor were consistently safer (consistency was statistically significant) than the shortest paths. The strategic paths at Grid level were safest among the three paths. We observed that the Grid paths were of shorter length, but in *in-game trials* it took more time to trace these paths compared to the Strategic paths. Also, the effect of variation in *checkpoint length* showed that for a shorter *checkpoint length* the *hit probability* computed for Meta-Weights was closer to the paths' evaluated *hit probability*.

5.1 Out-Game Trials for Paths

The hypothesis of this experiment was that the difference in hit probability between the shortest paths and the strategic paths is statistically significant. We computed strategic paths for 50 random experiments with one enemy. In these experiments a start area, an enemy area and a goal area were randomly selected from the available open areas. The strategic paths were computed for *RiskVsTime*=10 for an enemy with an assault rifle. We compared the shortest path and the strategic path at Area level using a t-test and found that the difference in path safety was statistically significant at the p=0.0056 level. Between the shortest path and the strategic path at Grid level the difference in path safety was statistically significant at the p= 0.00076 level. Between strategic paths at Area level and Grid level we found the difference in path safety was statistically significant at the p=0.00085 level. This confirms the claim that when seen in abstraction, an Area gives a rough estimation about its safety. And when that Area is reached and the Grid Path is then computed for that Area, this Grid Path can be consistently traversed with same or lesser risk.

5.2 Out-Game Trials for HP Variation

The hypothesis of this experiment was that with an increase in the *RiskVsTime* factor the generated strategic paths will become safer. Also with a decrease in the checkpoint length we will see a smaller difference between *hit probabilities (HP)* computed by the *checkpoint technique* for *Meta-Weight computation* and *risk evaluation* of the traced path.

Figure 4 shows that as *RiskVsTime* factor increases the *hit probability (HP)* decreases and thereby the computed path becomes safer. Also, Grid Paths are consistently safer compared to other paths. The *RiskVsTime* factor has no effect on the shortest path. When checkpoints are of smaller lengths the difference between *HP* for Meta-Weight and the Path's risk evaluation significantly decreases, and they tend to approach higher values compared to shorter checkpoint lengths.

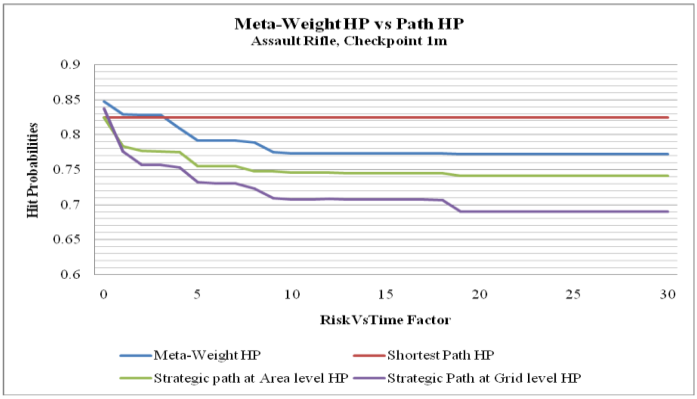


Fig. 4. HP variation between Meta-Weight computation and path's risk evaluation for 1m checkpoint length

5.3 Out-Game Trials for Path Distance

The hypothesis of this experiment was that with an increase in the *RiskVsTime* factor the distance of traversal will increase. In figure 5, for *out-game* trials as the *RiskVsTime* factor increases the strategic paths become safer at a cost of longer distances. The strategic path computation selects the shortest penalized path and in this process it tends to minimize both the risk and the distance of traversal.

As shown in figure 1 (Right) and figure 5, in the case of the strategic path at Area level the distance is the shortest distance between the *gateway points* lying at the centers of *Gateways*. And in the case of the strategic path at Grid level, these *gateway points* tend toward the goal area and strive to remain smooth over the irregular areas (using a technique similar to A*). As a result the distance is further minimized, possibly even below the shortest path length.

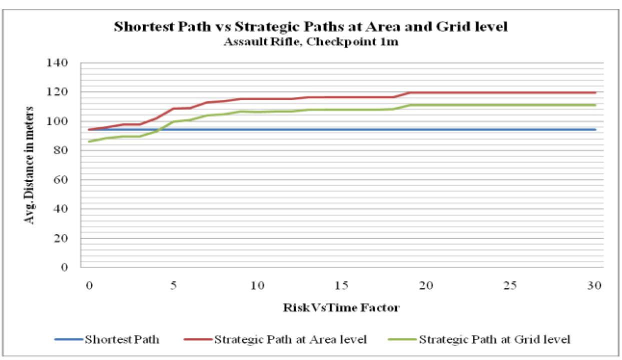


Fig. 5. Avg. distance of shortest paths and strategic paths over Area and Grid level

5.4 In-Game Trials

The hypothesis of this experiment was that the difference between the shortest paths and the strategic paths, as measured within the UCT game, is statistically significant.

For *RiskVsTime* = 10 we ran the same set of 50 experiments and compared the shortest path and the strategic path at Area level using a t-test and found that the difference in path safety was statistically significant at the p=0.006 level for the in-game trials. Between the shortest path and the strategic path at Grid level the difference in path safety was statistically significant at the p=0.0001 level for the in-game trials. Between strategic paths at Area level and Grid level we found the difference in path safety was statistically significant at the p=0.0036 level for the in-game trials.

During the in-game trials the shooting accuracy is based upon the angling calculations (weapon and the target), and the movement accuracy is based upon the bounding box [8] calculations (between objects and the agent). The movement computation tries to minimize any collision with the walls and the objects. Thus, the *in-game trials* contain many details where any technical inaccuracy could have a negative impact on the results. On the other hand for the case of *out-game trials* these game details are abstracted and do not adversely affect the analysis.

The hypothesis of this experiment was confirmed. The importance of the *in-game trials* was to check the accuracy of the *strategic path computation model*, and we found the model was consistent with the *in-game trials*.

6 Conclusions

The overall goal of this work is to improve the realism of paths taken by players in an urban warfare game. We have developed techniques for constructing strategic paths that take into account the desired risk vs. time tradeoff to find safer paths based on a model of an enemy and their different weapons. This model allows the computation of the probability that the player will be hit by the enemy while traversing the path, and therefore allows the tradeoff between risk and time. This model was evaluated using simulated trials (out-game), and the results were verified through comparison with actual in-game trials using the Urban Combat Testbed, a modification of Quake3. Results show that the model allows the selection of significantly safer paths, and that the path hit probabilities for the out-game trials are similar to those observed for the in-game trials. Future directions for this work include the further refinement of the strategic path model, extension of the approach to other maps and other MOUT games and simulators, further automation of the mechanisms for decomposing a map into areas to support area-level and grid-level strategic path planning, and ultimately integration of these techniques into MOUT game players to improve performance and realism.

Acknowledgements

This material is based on research sponsored by DARPA under agreement number FA8750-05-2-0283. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government.

References

1. Ghost Recon tactical information, http://www.ick.bz/pdfs/GR2_weapons-2005-03-25_1019.pdf
2. Kamphuis, A., Overmars, M.: Tactical Path Finding in Urban Environments. In: First International Workshop on Crowd Simulation (2005)
3. Kondeti, B., Nallacharu, M., Youngblood, G.M., Holder, L.B.: Interfacing the D'Artagnan Cognitive Architecture to the Urban Terror First-Person Shooter Game. In: IJCAI Workshop on Reasoning, Representation and Learning in Computer Games (2005)
4. Liden, L.: Using Nodes to Develop Strategies for Combat with Multiple Enemies. In: AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment (2001)
5. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice-Hall, Englewood Cliffs (1994)
6. Singh, A.: Improving the Survivability of Agents in a First-Person Shooter Urban Combat Simulation by Incorporating Military Skills. Master of Science Thesis, Washington State Univ. (2007)
7. Urban Terror, http://www.urbanterror.net
8. Waveren, J.P.: The Quake III Arena Bot. Master of Science thesis, Delft University of Technology (2001)
9. Youngblood, G.M., Nolen, B., Ross, M., Holder, L.B.: Building Test Beds for AI with the Q3 Mod Base. In: Artificial Intelligence in Interactive Digital Entertainment Conference (2006)

Fear Inducer: A Mixed Reality Audio Experience

Jurgen Westerhoff

Eindhoven university of Technology, Eindhoven

Abstract. In this paper I present the results from a design research project done at the Eindhoven University of Technology. In this project I explored the possibilities of mixing realities using sound. With this mixed reality I explored ways to make people scared for entertainment purposes, comparable to horror-movies or thrill rides. One of the main elements of the concept was to make the device mobile, so that a thrilling experience can be generated at any time and any place. The user will put on the audio-device and through a biofeedback-loop connect himself to the system. While walking around sounds come from the system that blend in with sounds from the real environment. This effect causes interesting situations, directed by the users own impulses sensed by the biofeedback-loop.

Keywords: mixed reality, audio experience, thrilling experience, biofeedback-loop.

1 Introduction

Recent trends in horror cinema show that there is a need for more extreme content and experiences in certain audiences. This is probably due to saturation of the audience,[1] and a need for new and exciting experiences. In this project different ways to scare people for entertainment purposes were explored, and how this can be done most effectively. Using a 'research through design' approach the user-experience was investigated with simple prototyping to see how people react to 'scary' stimuli.

Mixed reality [2] systems offer ways to blend the real with the virtual, and hereby open up a lot of possibilities for entertainment products. Most of the mixed reality systems however are focussed on visual effects[3][4], but audio has much potential as well. In this project the use of mixed reality in audio is investigated through the design of a product for a specific target group.

A large part of the effects induced by scary movies is related to instinctive reactions in the brain [5]. Apparently, the acted events in a movie trigger instinctive reactions and put the viewer in a state of 'fight or flight' just as real life situations would. This heightened state of arousal has some measurable effects on the human body, such as a change in heart-rate and sweating.

2 Blending Real and Virtual

2.1 Letting the Imagination Run Wild

One of the reasons that horror movies are getting boring for audiences is that the location and predictability of the scripts are often the same. A movie fan will be able to

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 88–93, 2008.

recognize style-elements and will get used to the familiar location of the theatre. All these things prohibit the imagination of the audience to run wild, this usually happens only after the movie. The opportunities lie in this after-effect of the movie, what happens when you keep on giving scary stimuli in this phase? The combination of subtle stimuli and the imagination of the audience could lead to tailored and non-scripted experiences.

This is where mixed reality starts to become interesting. In stead of the fixed setting of the theatre, there are now possibilities to combine virtual information with real world settings. The challenge is to feed the imagination with stimuli, and take away the boundary between what is real and what is virtual.

2.2 Auditive Stimuli

Auditive stimuli are very effective in inducing fear [5]. They affect a part of the brain that is quickly startled. This is an evolutionary warning system, used to alert you when unknown sounds appear. This is especially true when the context is unknown and the sounds are not easily recognized. Because of this primal reaction, audio is very suited to make people scared. This is also the reason why scary movies often have a very heavy sound-score.

Another plus-side of using auditive stimuli is that they blend seamlessly with real-world sounds (depending on the type of headphones). It keeps the field of view unhindered, and leaves room for imagination. This makes it ideal for a product in which freedom of movement is needed.

2.3 Biofeedback

Blending the virtual sounds with the real world poses some challenges. One of which is how to now when a person reacts to the sounds in the environment. With visual stimuli it is possible to track where a person is looking, and how he manipulates the real world. With sound however it is much harder to know which sounds have an effect on someone. To get an idea about which sounds have an effect on a person, I introduced a biofeedback system. This system uses galvanic skin response to sense the stress-levels of the user. By doing this the system is able to react on objects in the surroundings of the user, and adjusting to that. So if for instance the user is walking past a dark corridor and the stress levels rise, the system is able to activate a scary sound to react to the reaction of the user. This creates a loop in which the experience of the user activates virtual sounds which in their turn influence the perception of the real world.

This use of biofeedback creates the link between the virtual sounds and the stimuli coming from the real world. The resulting effect is that in the perception of the user the sounds coming from the system and the sounds from the real world blend together.

2.4 The Scare Effect

This blending of real and virtual is what makes the experience scary. During the user tests it became apparent that this state (of blending) is not immediately in effect. At first the difference between the virtual and the real world is distinguishable, but after a while the devision starts to become vague. Especially abstract sounds are hard to pinpoint. Things like the wind or a barking dog are hard to place in either the virtual or real world. The resulting effect is a state in which the user does not know where sounds are coming from, and this gives a very menacing and scary experience. You can imagine for instance that when a virtual sound of a barking dog starts coming towards you, you cannot help but look behind you to check if this is not happening in the real world. This effect intensifies over time while using the system. Test-subjects expressed that the final minutes of the test were scariest because they had become paranoid of all the sounds that came from the system.

3 The Fear Inducer System

To test the effects of the audio-experience a fully working prototype was made. This prototype consisted out of three elements; a headset, a biofeedback loop and a series of audio-samples. The user puts the headset on, and walks around outside. Fig.1 shows a short scenario of how the system could be used. The system is able to detect rises in stress-levels, and will play a scary audio sample as a reaction to that. The prototype was tested on a small group of test-subjects, who had to walk a given route in a context unknown to them. This test was highly experimental, and only gave an impression on how the subjects experienced the prototype. The results from this test showed that the first impressions of the subjects was very positive, they experienced it as very thrilling and were willing to try it again.

I will shortly describe the workings of the separate elements, and how they affect the overall experience of using the prototype.

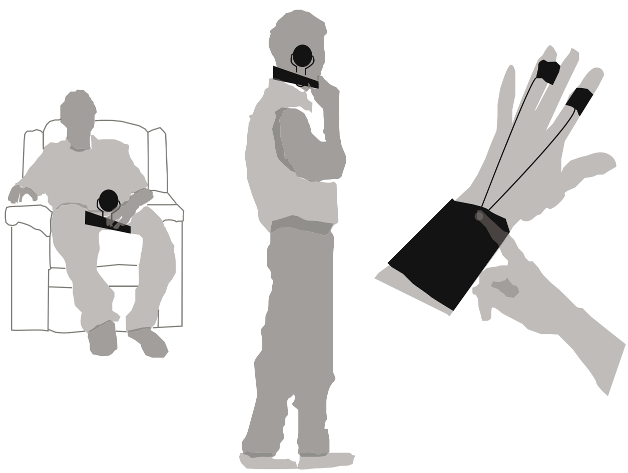


Fig. 1. Putting on the headset and biofeedback system

3.1 Design of the Headset

Next to the auditive stimuli, the headset itself also contributes to the experience. Throughout a series of pilot tests it became clear that not being able to look behind you has a big effect on how scary people perceived a context. This passive haptic feedback does nothing more than restrict head-movements. The effect however on the overall experience is considerable. Next to this restriction the overall design of the headset is quite menacing, and looks unpleasant to put on (actually wearing the headset however is quite comfortable). Fig.2 shows a picture of the final prototype headset. These design-choices came directly from several levels of user involvement throughout the project. The overall look and feel were inspired by a persona [7] that was made about the targeted user group. The challenge was to make it fit with other products used by the target group, without having to lean to much on classical horror-themes.

3.2 The Biofeedback Loop

The biofeedback system works by measuring the galvanic skin response(gsr) of the user. When the user is in a stressful state, the system will be able to read this and give back a value. It basically is a very simple electronic circuit in which the user is used as a variable resistor. In the prototype the gsr-measurement is done at the fingertips, because this gives to clearest reading.

To explore how much the body reacts to scary stimuli, a small test was performed. Several test-subjects were placed in front of a screen and attached to a biofeedback measurement circuit (measuring gsr). On the screen a scene from a recent horror-movie was shown, and the gsr-readings were recorded on a computer. At scary parts in the movie the gsr-reading clearly showed peaks, about three seconds after the actual scary scene. This delay is probably a sum of the lag in sweat building up, and the processing of the sensor-reading. Because the readings gave clear enough results to detect peaks in scariness, the system proved itself robust enough for use in a prototype.

Fig. 2. The final prototype headset

3.3 The Audio

In the prototype the audio was stored and played on two separate mp3-players. These players were connected to the biofeedback system, and would play a song when a certain threshold in the biofeedback-readings was reached. The audio samplers were divided in two types, one mp3-player held the ambient sounds and the other one the audio-effects. By using two players, the audio was mixed seamlessly.

The ambient sounds would play continuously, and increased in volume and intensity during the length of the test. This created a suspense, not unlike what is used in movies[8]. The sound effects however were only played when a peak in the biofeedback readings was measured (what meant that the test-subject was excited by something). The sound effects ranged from barking dog sounds to more abstract howling sounds. During the user-test the subjects expressed that the more abstract sounds that seemed to be coming towards them (with virtual 3d sound techniques) were scariest.

4 Conclusion

Through this paper I attempted to describe the progress I made with showing the possibilities of mixed reality using audio. Although the project was highly experimental, I hope to show the potential of mixed reality in audio through the form of a product. One of the advantages of the 'research through design' approach is that you can take the wishes and reactions of real people into the process. The reactions of people to the final prototype, and the involvement of users in early explorations in the project helped the development take big steps in a short amount of time. With this paper I want to show an interesting direction of design research, and hope to encourage people to develop more entertainment devices using the possibilities of mixed reality.

References

1. Rauterberg, M.: About a framework for information and information processing of learning systems. In: Falkenberg, E., Hesse, W., Olive, A. (eds.) Information System Concepts – Towards a consolidation of views, pp. 54–69 (1995)
2. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IECE Transactions on Information Systems E77-D(12) (December 1994)
3. Cho, K., Kang, W., Soh, J., Lee, J., Yang, H.S., Hunter, G.: A handheld augmented reality game system with dynamic environment. In: Ma, L., Rauterberg, M., Nakatsu, R. (eds.) ICEC 2007. LNCS, vol. 4740, pp. 10–15. Springer, Heidelberg (2007)
4. Abawi, D.F., Reinhold, S., Dorner, R.: A Toolkit for authoring non-linear storytelling environments using mixed reality. In: Göbel, S., Spierling, U., Hoffmann, A., Iurgel, I., Schneider, O., Dechau, J., Feix, A. (eds.) TIDSE 2004. LNCS, vol. 3105, pp. 113–118. Springer, Heidelberg (2004)
5. Thompson, B.: Evoking Terror In Film Scores. M/C: A Journal of Media and Culture 5.1 (2002), http://www.media-culture.org.au/0203/evoking.php

6. LeDoux, J.: The Emotional Brain: The Mysterious Underpinnings of Emotional Life. Touchstone, New York (1996)
7. Olsen, G.: Persona Creation and Toolkit (2004), `http://www.interactiondesigners.com`
8. Sirish, K.A.: Computational model of suspense in virtual worlds, Raleigh (2003)

Game Bot Detection Based on Avatar Trajectory*

Kuan-Ta Chen[1], Andrew Liao[2], Hsing-Kuo Kenneth Pao[3], and Hao-Hua Chu[2]

[1] Institute of Information Science, Academia Sinica
[2] Dept. of Computer Science & Information Engineering, National Taiwan University
[3] Dept. of Computer Science & Information Engineering,
National Taiwan Univ. of Science & Technology

Abstract. In recent years, online gaming has become one of the most popular Internet activities, but cheating activity, such as the use of game bots, has increased as a consequence. Generally, the gaming community disagrees with the use of game bots, as bot users obtain unreasonable rewards without corresponding efforts. However, bots are hard to detect because they are designed to simulate human game playing behavior and they follow game rules exactly. Existing detection approaches either interrupt the players' gaming experience, or they assume game bots are run as standalone clients or assigned a specific goal, such as aim bots in FPS games.

In this paper, we propose a trajectory-based approach to detect game bots. It is a general technique that can be applied to any game in which the avatar's movement is controlled directly by the players. Through real-life data traces, we show that the trajectories of human players and those of game bots are very different. In addition, although game bots may endeavor to simulate players' decisions, certain human behavior patterns are difficult to mimic because they are AI-hard. Taking Quake 2 as a case study, we evaluate our scheme's performance based on real-life traces. The results show that the scheme can achieve a detection accuracy of 95% or higher given a trace of 200 seconds or longer.

Keywords: Cheating Detection, Online Games, Quake, Security, Supervised Classification, User Behavior.

1 Introduction

In recent years, online gaming has become one of the most popular Internet activities. However, as the population of online gamers has increased, game cheating problems, such as the use of *game bots*, have become more serious. Game bots

* This work was supported in part by Taiwan Information Security Center (TWISC), National Science Council under the grants NSC 97-2219-E-001-001 and NSC 97-2219-E-011-006. It was also supported in part by Taiwan E-Learning & Digital Archives Program (TELDAP), National Science Council under the grants NSC 96-3113-H-001-010 and NSC 96-3113-H-001-012.

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 94–105, 2008.

are automated programs with artificial intelligence that players use for different purposes. In MMORPGs (Massively Multiplayer Online Role Player Games), players can save a great deal of time by using bots to perform repetitive tasks, such as slashing low-level monsters, or fishing in a river to master the avatar's fishing skills. In FPS (First-Person Shooter) games, users can employ bots to play in place of themselves in order to get high scores and gain a reputation in the community.

Generally, the gaming community disagrees with the use of game bots, as bot users obtain unreasonable rewards without corresponding efforts. However, game bots are hard to detect because *they are designed to simulate human game playing behavior and they follow game rules exactly*. Some bot detection studies [1, 2] propose using CAPTCHA tests during a game to determine whether an avatar is actually controlled by a person. Although this method is effective, it interrupts the game play and degrades players' feelings of immersion in the virtual world [3,4]. Alternatively, passive detection approaches, such as schemes based on traffic analysis [5,6] and schemes based on avatars' shooting accuracy in FPS games [7], can be used. The former approach assumes that a game bot works as a standalone client, and the latter is only valid for detecting aim bots in shooting games.

In this paper, we propose a general approach for *all genres of games where players control the avatar's movement directly*. Our approach is based on the *avatar's movement trajectory* during a game. The rationale is that *the trajectory of the avatar controlled by a human player is hard to simulate*. Players control the movement of avatars based on their knowledge, experience, intuition, and a great deal of information provided in the game. Since human decisions may not always be logical and efficient, how to model and simulate realistic movements is still an open question in the AI field. To distinguish human players from game bots efficiently, we analyze the trajectories of both player types and distinguish between the trajectories according to their *spatial and temporal characteristics*. We choose Quake 2 as our case study because it is a classic and popular FPS game, and many real-life human traces are available on the Internet. Therefore, we can use such traces to validate our proposed scheme.

The contribution of this paper is two-fold. 1) We propose a trajectory-based approach for detecting game bots. It is a general model that can be applied to any game in which the avatar's movement is controlled by the players directly. 2) Using real-life human traces, the performance evaluation results show that the scheme can achieve a detection accuracy of 95% or higher when the trace length is 200 seconds or longer. Because it is difficult to simulate human players' logic when they control game characters, we believe this approach has the potential to distinguish between human players and automated programs and thus merits further investigation.

The remainder of this paper is organized as follows. Section 2 contains a review of related works. In Section 3, we introduce our game case study, Quake 2, and describe the game trace collection methodology. We analyze the similarities and differences between the trajectories of different types of players in Section 4. In

Section 5, we propose an identification scheme and demonstrate its ability in terms of the distribution of discriminative features. In Section 6, we evaluate the performance of the proposed scheme with the consideration of the trace length. Then, in Section 7, we summarize our conclusions.

2 Related Work

Recently, anti-cheating software programs, such as PunkBuster and GameGuard, have been widely deployed in online games to prevent cheating. Such software is bundled with game clients, so it cannot be uninstalled even if the game client has been removed. It works by hiding in the game client process, monitoring the entire virtual memory space (to prevent modification of the game's executable images), blocking suspected programs that might be hacker tools, and blocking certain API calls. This kind of software can detect nearly all plug-in tools that attach to a game client program to inspect or modify game states when the game is running. Unfortunately, it cannot stop the widespread use of standalone bots, including the bot series we study in this paper. The reason is that these anti-cheating software programs are host-based, so they must be installed on players' PCs to be effective. Standalone bots, on the other hand, can function *without* clients, and it is unlikely that anti-cheating tools would be installed on PCs where the bots are running. This claim is strongly supported by the fact that game bots are still active in games protected by PunkBuster or GameGuard, e.g., Quake (PunkBuster) and Lineage[1] (GameGuard).

3 Data Description

3.1 Human Traces

Quake 2 supports a game-play recording function that keeps track of every action and movement, as well as the status of each character and item throughout the game. With a recorded trace, one can reconstruct a game and review it from any position and angle desired via VCR-like operations. Players often use this function to assess their performance and combat strategies. Moreover, experienced players are encouraged to publish their game-play traces as teaching materials for novice gamers and thereby build a reputation in the community.

To ensure that our game traces represented the diversity of Quake players, we only used traces that players had contributed voluntarily. The human players' traces were downloaded from the following archive sites: GotFrag Quake[2], Planet Quake[3], Demo Squad[4], and Revilla Quake Site[5]. We restricted the traces to the map *The Edge*, one of the most well-known levels of death-match play. On

[1] `http://boards.lineage2.com/showflat.php?Number=573737`

[2] `http://www.gotfrag.com/quake/home/`

[3] `http://planetquake.gamespy.com/`

[4] `http://q2scene.net/ds/`

[5] `http://www.revilla.nildram.co.uk/demos-full.htm`

Table 1. Trace Summary

	name	num	mean	total	active
1	Human	93	2 hour	203.5 hour	91%
2	CR	24	19 hour	448.8 hour	91%
3	Eraser	15	20 hour	296.4 hour	94%
4	ICE	18	20 hour	358.8 hour	67%

this map, the only goal is that each player should kill as many other players as possible, until the time limit is reached. Because short traces contain little information, we only collected traces longer than 600 seconds.

3.2 Bot Traces

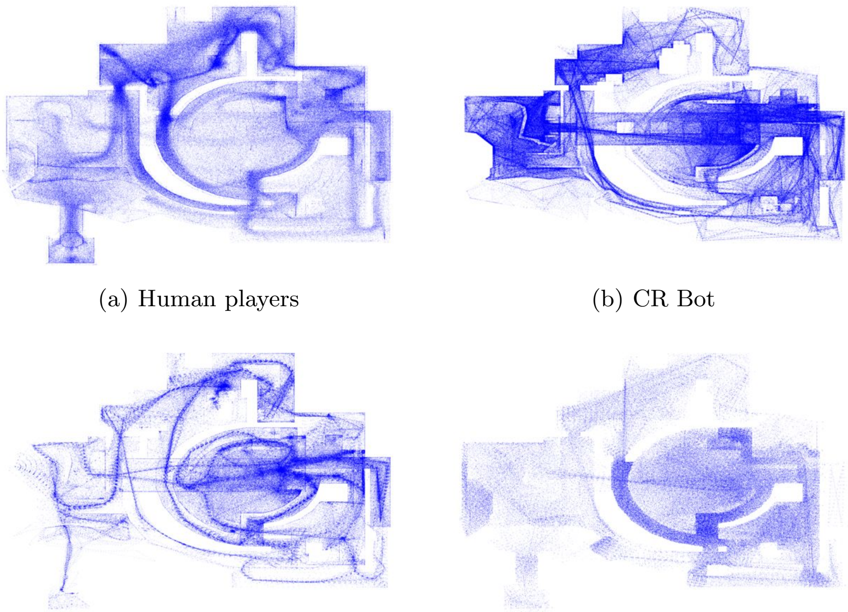
There are many game bots available for Quake 2. For this study, we selected three of the most popular bot programs for trace collection, namely CR BOT 1.14 [8], Eraser Bot 1.01 [9], ICE Bot 1.0 [10]. We collected $1,306$ hours of traces in total, as shown in Table 1. In CR Bot and Eraser Bot, all human players and bots were active most of time ($\geq 90\%$). There was less activity in ICE Bot because it often remained idle in some places waiting for an opportunity to ambush other players.

4 Discriminative Analysis

In this section, we compare the avatar trajectories of human players and game bots. First, we compare the navigation patterns of the two player types and consider their individual trajectories. We then identify the most significant discriminative characteristics of the respective trajectories and incorporate them into the proposed bot identification scheme.

We construct the aggregated navigation pattern of each player type by plotting all the observed coordinates in all traces of the particular player type on a graph, as shown in Fig. 1. The areas of high density in each figure are the places that players visit more frequently, while the sparse areas represent buildings or other types of obstacles that players cannot pass. The figures show that the game level is formed by squares, plazas, and narrow corridors. This arrangement is designed specifically for death-match play, as the winding routes provide cover for players to hide, and the narrow corridors lead to intense fighting if players confront each other in these places. We observe that, even though all the movement traces were collected on the same map, the navigation patterns of different player types are dissimilar. We summarize the differences below.

1. Human players tended to explore all areas on the map; thus, Fig. 1(a) shows the most complete terrain of the level. In contrast, the routing algorithms of game bots may have had difficulty navigating to some places, so they never visited some parts of the map. For example, the bottom left-hand corner of the CR Bot navigation map in Fig. 1(b) does not indicate the presence of bots.



(a) Human players (b) CR Bot

(c) Eraser Bot (d) ICE Bot

Fig. 1. Presence locations of all players

2. To reduce the probability of being attacked, human players normally avoid open spaces. Therefore, in Fig. 1(a) we observe that human players avoided the plaza in the middle of the map, and stayed in the surrounding corridors instead. This is indicated by the high density of plots in the corridors. In contrast, game bots often stay in the central plaza, probably because it occupies a large space and it is easy to get everywhere from this area.
3. Even though human players spend most of their time in narrow areas and confined rooms, there are large variations in their trajectories. There are two reasons for this phenomenon. 1) The width of the main routes is quite large. Rather than stay in the middle of a route, players move irregularly within the limited space. This may be due to players' preferences; hence, some players may move along the wall of the path, while others may walk straight, unless the avatar is blocked by a wall or other obstacles. 2) As fights may occur anytime-anywhere, human players often move strategically to dodge current or potential attacks. On the other hand, we find that different game bots adopt very different movement patterns over the routes. The movement paths of CR Bot and Eraser Bot (Fig. 1(b) and Fig. 1(c) respectively) are dense and easy to see. This suggests that these bots tend to follow exact movement patterns when moving through the same corridor. In contrast, ICE Bot (Fig. 1(d)) exhibits a nearly uniform distribution over all possible points on the map. This implies that its routing algorithm decides the avatar's

direction rather than its exact movement pattern, so that the probabilities of all points on the route are roughly equivalent.

Clearly, there are substantial differences between the aggregated navigation patterns of human players and those of each game bot because the bots' routing patterns are very different from the movement behavior exhibited by human players.

5 Bot Detection Scheme

Our objective is to classify human players and game bots efficiently and accurately. To this end, we integrate the spatial and temporal differences in the trajectories of avatars controlled by different player types to develop a bot identification scheme. In this section, we first describe the set of discriminative features extracted from the avatar trajectories, and then explain how we use the features to classify game bots and human players.

5.1 Feature Extraction

Given a segment of a trajectory, $\{x_t, y_t\}, 1 \leq t \leq T$, we extract the following features from this two-dimensional time series.

1. ON/OFF Activity. First, we note that avatars in the game play do not move all the time. Sometimes they may stop to check if any opponents are around, wait for opponents to enter an area, wait for regeneration of their weapons or ammunition, or simply take a rest. The alternate moving and idle behavior forms an ON/OFF movement pattern. We define ON periods as consecutive periods of movement longer than 1 second, and OFF periods as the remaining time frames. The duration and frequency of ON/OFF periods are decided by the players' styles and the bots' AI logic. For example, aggressive players may keep moving all the time, while cautious players may stay in one place to monitor their surroundings. Therefore, we define four features based on ON/OFF activity: *the mean and standard deviation of ON periods, and those of OFF periods.*

Fig. 2 shows the distributions of the four features. The mean and standard deviation of human players' ON periods are significantly higher than those of game bots. This indicates that human players are more aggressive as they tend to move all the time. In addition, the mean and standard deviation of human players' OFF periods are longer than those of bots, which implies that human behavior is more irregular and unpredictable in that they may wait for a longer time after a long move. The figure shows that human players and game bots differ in terms of ON/OFF activity. Hence, we believe that the four features based on these activities could be useful for bot detection.

2. Pace. In games, avatars are generally allowed to move at different speeds and in different ways, such as running, slow walking, step-by-step walking, lateral

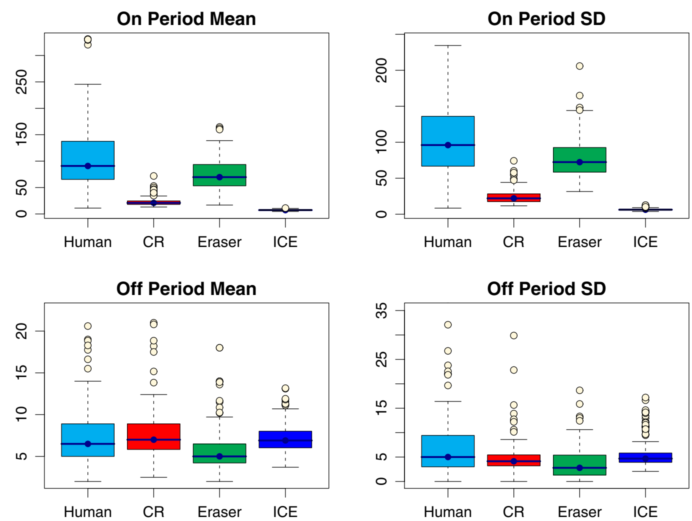


Fig. 2. The distribution of features related to ON/OFF periods

shifting, and moving backwards. In addition, players can stop the current movement and proceed with another movement in different direction in sub-seconds; therefore, the resulting avatar movements can be highly variable. One simple way to characterize the dynamics of an avatar's movement is by the pace of its movements. We define the *pace* as the displacement of an avatar's coordinate in one second, and extract *the mean and standard deviation of the pace* as two features. We find that the paces of most avatars are generally small, although they can be large occasionally. To characterize the variability of paces when players move fast, we also define *the "large pace SD," which is the standard deviation of paces larger than* 10 *units*.

In addition to normal movements, players may teleport their avatars to a remote place instantly through a teleportation spot. Teleportation may also be used when an avatar dies. It will be transferred to the rebirth spot so that its life points can be recharged. We detect teleportation occurrences by computing if the offset in one second is longer than 60 units and define the feature "teleportation rate" as *the average count of teleportation occurrences recorded in one second*.

Fig. 3 shows the distribution of the four features related to the movement pace and teleportation. Although the means of the paces of different player types are dissimilar, the variations are not large. This shows that the four player types have different but consistent micro-movement behavior in small time scales. The standard deviation of the pace also has large discriminability, where that of human players and Eraser Bot have similar magnitude. The large standard deviation of the pace, on the other hand, exhibits great discriminability, which indicates that human players have even larger pace variability when they move fast. Finally, CR Bot and Eraser Bot have very low teleportation frequency. In contrast, human players have moderate teleportation frequency. Moreover, their

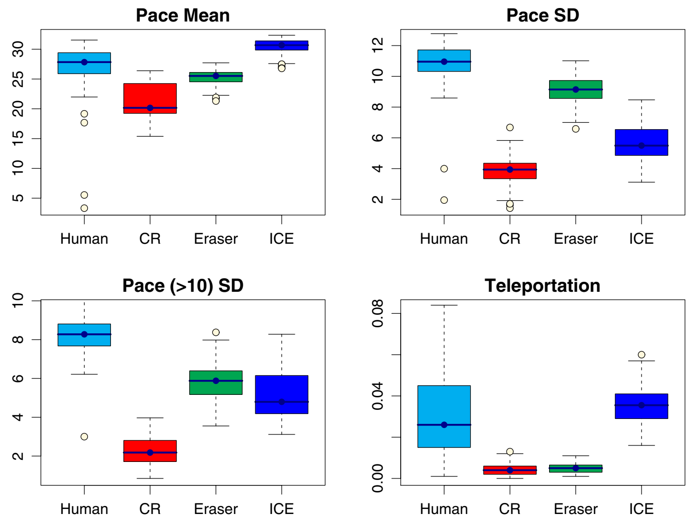


Fig. 3. The distribution of features related to movement pace

variance is high because human players have different preferences when using teleportation spots and players get killed at different rates. .

3. Path We also define the following features to characterize the detailed trajectories of avatars in a game.

Lingering. We consider whether players "lingered" in a small area during a specific time period. For an avatar at (x, y) at time t, if its distance from (x, y) was always less than d during the period $(t, t+p)$, we say that the avatar was lingering during $(t, t+p)$, given the parameters (d, p). We arbitrarily set $d = 30$ seconds and $p = 300$ units, as we find that different parameters do not affect the classification performance significantly.

Smoothness. The "smoothness" feature determines whether an avatar moves in straight or zig-zag patterns. Assume an avatar is at (x_1, y_1) at time t_1 and at (x_2, y_2) at time t_2. We define the smoothness as the number of times the character moves across the line $(x_1, y_1) - (x_2, y_2)$ during the period (t_1, t_2). As the line $(x_1, y_1) - (x_2, y_2)$ indicates the shortest route between the two points (x_1, y_1) and (x_2, y_2), crossing the line implies that the player is moving inefficiently. This may be because he is attempting to dodge gunfire, switch to another target, or simply due to players' habits or bots' routing algorithms.

Detour. We define another feature "detour" to quantify the effectiveness of user movements. If an avatar is at (x_1, y_1) at time t_1 and at (x_2, y_2) at time t_2, we compute the detour by dividing the length of the movement by the effective offset of an avatar during the period (t_1, t_2).

The distributions of the above features are plotted in Fig. 4. The graph shows that the linger frequency and duration of human players are significantly less

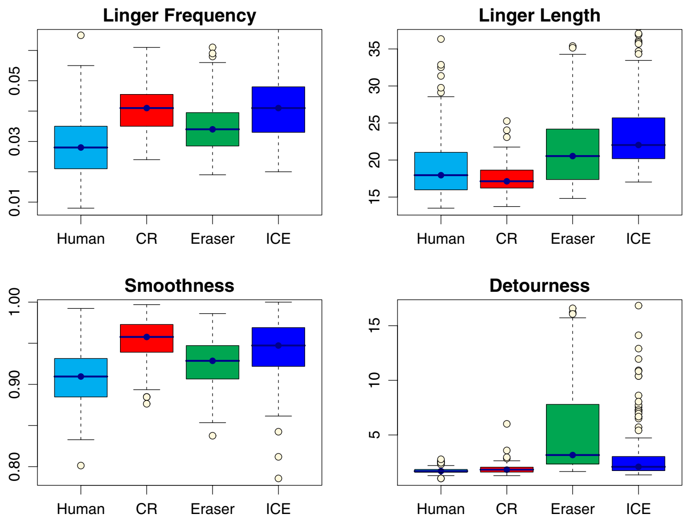


Fig. 4. The distribution of features related to movement path

than those of game bots. This is reasonable because lingering in a place for a long time is a dangerous, as the player may be noticed and induce opponents' fire. The smoothness of human players is the lowest of the four player types, which supports the intuition that human players' movements are the most irregular and unpredictable. The detour feature shows that Eraser Bot moves very inefficiently in terms of the avatar's effective offset. In contrast, the movements of human players are relatively more efficient. We suspect this is because *human players tend to move away from current positions to another place efficiently even though they may move irregularly and strategically; thus, the resulting avatar trajectory exhibits both unpredictability and efficiency which seem contradictory.*

4. Turn. Our final set of features is based on the frequency and amplitude of how avatars change direction. Our rationale is that each time an avatar changes direction, the magnitude of the change should be dependent on player conventions and bot routing algorithms.

Assume an avatar is at (x_1, y_1) at time t, at (x_2, y_2) at time $t+p$, and at (x_3, y_3) at time $t + 2p$. If the angle between two vectors $(x_2 - x_1, y_2 - y_1)$ and $(x_3 - x_1, y_3 - y_1)$ is greater than a, we determine that a turn with angle a occurred. We define three features to denote the frequency of turns with angles 30°, 60°, and 90°, respectively. In addition, we define a feature called the "turn angle" to denote the average angle change for all direction changes greater than 30°.

Fig. 5 shows the distributions of the turn-related features. We observe that the four player types change direction at different rates no matter how we define the minimum degree of a direction change. Notably, the turn frequency of human players is the highest for the 30° angle and becomes relatively lower for the 90° angle. In addition, the average turn angle of human players is the lowest among

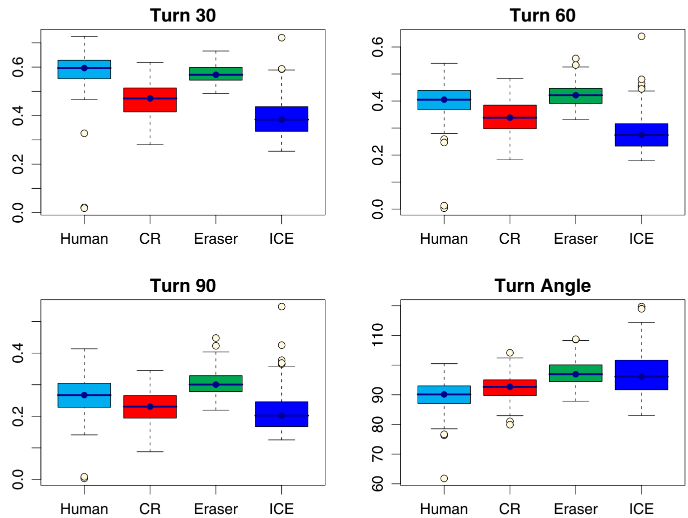


Fig. 5. The distribution of features related to turn movement

the four types, which indicates that *human players tend to adjust their directions continuously and slightly.*

5.2 Classification

We apply a supervised classification framework to train a classifier, which we use to determine whether a segment of an avatar's trajectory belongs to a human player or a game bot. The classifier we adopt is the naive Bayesian classifier without the kernel density estimation. We evaluate the performance of trajectory classification in the next section.

6 Performance Evaluation

In this section, we evaluate the performance of our proposed bot detection scheme on the collected traces. First, we evaluate whether our scheme can distinguish between human players and game bots, by using the classifier to perform 10-fold cross-validation. In real-life scenarios, the trace length plays an important role because it determines how quickly a game bot can be detected. Thus, we evaluated the performance of our scheme on different traces lengths, as shown in Fig. 6. The graph shows that the detection accuracy is higher than 90%, even when the trace length is as short as 100 seconds. Longer traces yield better accuracy. To determine which category of features yields the highest accuracy, we plot the classification performance for each category of features. The results indicate that the features related to the movement pace, direction changes, and ON/OFF periods all yield good results, while path-related features only exhibit good discriminability when the trace length is 800 seconds or longer.

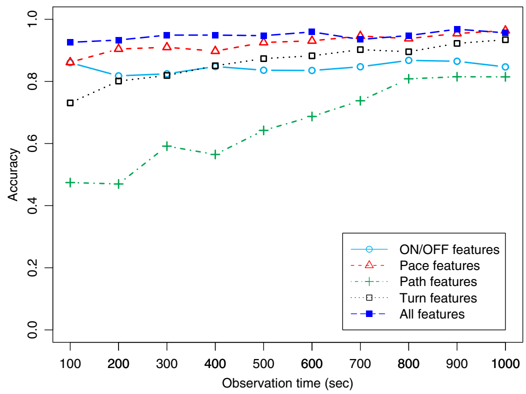


Fig. 6. Accuracy between human and bots

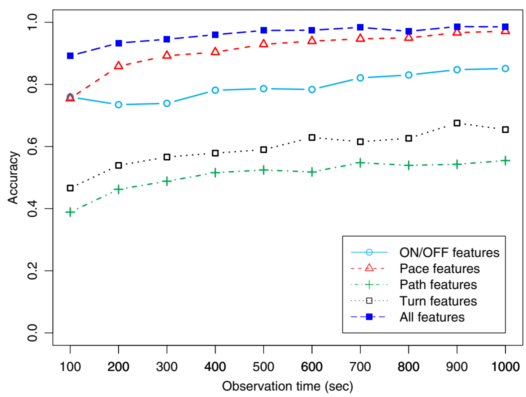


Fig. 7. Classification accuracy between four types of players (human and three bot programs)

Furthermore, we perform a player-type classification; that is, we not only determine whether a character is controlled by a human player or a bot program, but also identify the bot program used if appropriate. The results are shown in Fig. 7. The classification accuracy of the player types is even better than that of the human-bot scenario when the trace length is longer than 200 seconds. With a trace length of 500 seconds or longer, our scheme yields a classification accuracy of 98% or higher. However, in this setting, individual feature categories, except those related to movement paces, exhibit low discriminability when they are applied to the classification separately.

7 Conclusion

We have proposed a trajectory-based approach for detecting game bots. It is a general technique that can be applied to any game in which the avatar's movement is controlled by the players directly. Our analysis of real-life traces shows that the trajectories of human players and game bots are very dissimilar. The performance evaluation results show that our bot detection scheme can achieve a detection accuracy of 95% or higher when the trace length is 200 seconds or longer. Because it is difficult to simulate human players' behavior when controlling game characters, we believe our method has the potential to distinguish between human players and automated programs, and thus merits further investigation.

References

1. Golle, P., Ducheneaut, N.: Preventing bots from playing online games. Computers in Entertainment 3(3), 3 (2005)
2. von Ahn, L., Blum, M., Hopper, N.J., Langford, J.: CAPTCHA: Using hard AI problems for security. In: Proceedings of Eurocrypt, pp. 294–311 (2003)
3. Novak, T.P., Hoffman, D.L., Duhachek, A.: The influence of goal-directed and experiential activities on online flow experiences. Journal of Consumer Psychology 13(1), 3–16 (2003)
4. Ila, S., Mizerski, D., Lam, D.: Comparing the effect of habit in the online game play of australian and indonesian gamers. In: Proceedings of the Australia and New Zealand Marketing Association Conference (2003)
5. Chen, K.T., Jiang, J.W., Huang, P., Chu, H.H., Lei, C.L., Chen, W.C.: Identifying MMORPG bots: A traffic analysis approach. In: Proceedings of ACM SIGCHI ACE 2006, Los Angeles, USA (June 2006)
6. Chen, K.T., Huang, P., Lei, C.L.: Game traffic analysis: An MMORPG perspective. Computer Networks 50(16), 3002–3023 (2006)
7. Yeung, S., Lui, J., Liu, J., Yan, J.: Detecting cheaters for multiplayer games: theory, design and implementation. Proc IEEE CCNC 6, 1178–1182
8. Malakhov, M.: CR Bot 1.15 (May 2000), http://arton.cunst.net/quake/crbot/
9. Feltrin, R.R.: Eraser Bot 1.01 (May 2000), http://downloads.gamezone.com/demos/d9862.htm
10. jibe: ICE Bot 1.0 (1998), http://ice.planetquake.gamespy.com/

Achievement of Carrying Objects by Small-Sized Humanoid Robot

Hiroyuki Nakagawa[1] and Ryohei Nakatsu[2]

[1] Kwansei Gakuin University, School of Science and Technology
2-1 Gakuen, Sanda, Hyogo, 669-1337 Japan
nakagawa-hiroyuki@kwansei.ac.jp
[2] National University of Singapore, Interactive & Digital Media Institute
Blk E3A #02-04, 7 Engineering Drive 1, Singapore 117574
idmdir@nus.edu.sg

Abstract. It is expected that the sphere of activity of robots will spread widely in the near future not only in fields of industry but to various aspects of our everyday life. On the basis of this expectation, in this research we aimed at the creation of a robot that supports our daily life and tried to develop a small-sized humanoid robot that can find an object, pick it up, and carry it. We developed an algorithm for a robot to recognize an object using color information. Also, based on the algorithm the robot calculates the direction and the distance to the object, moves to the precise position and picks up the object and carries it. Finally, we carried out an experiment to evaluate the performance and were able to obtain a high success rate.

Keywords: humanoid robot, object recognition, physical support.

1 Introduction

Various types of robots are used mainly in industrial areas in modern society. It is expected that new types of robots such as humanoid robots and pet robots will emerge and be introduced in our society in the near future. Corresponding to this expectation, research and development of humanoid and pet robots is flourishing [1][2]. So far the major application area for most of these robots is entertainment, such as playing soccer games [3], dancing [4][5], and so on. At the same time, however, it is expected that in the future these robots will support the physical aspects of our life like other industrial robots. Therefore, it is necessary to carry out a study on allowing these robots to help us in our daily life. In this study, we aim at achieving a function of carrying an object for a small-sized humanoid robot. For this purpose it is necessary for the humanoid robot to have capabilities of recognizing an object, moving toward it, picking it up, and carrying it. Based on this consideration, we developed an algorithm for a humanoid robot to fulfill the above functions. In the following, details of the algorithm are described. Also, we carried out experiments to evaluate the algorithm we developed, and the details of the evaluation experiments will be described.

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 106–115, 2008.

2 Related Studies and the Aim of the Research

Various types of pet robots and humanoid robots have been studied and developed. One of the most famous pet robots is a dog-like robot called AIBO [1]. Also, there are several humanoid robots being studied and developed. One of the representative humanoid robots is ASIMO [2]. The main role of AIBO is entertaining people by achieving various types of dog-like behaviors. On the other hand, it is expected that ASIMO would help us in many ways in our daily life, though so far, the main role of ASIMO is to perform various types of human-like actions such as walking and going up/down stairs at events and thus entertaining us. On the other hand, there are several studies underway that aim at supporting our life physically using humanoid robots. One of the studies, using a humanoid robot called HRP2, has examined how to achieve life support functions such as dish washing or room cleaning [6]. In other research, an assistant robot in a nursing home has been studied [7]. This robot has the capability of finding a person, giving a reminder to a person, accompanying a person to a desired place, giving a person required information, and so on.

In the case of a small-sized humanoid robot, as it becomes much harder for these robots to support us physically in our daily life, most of the research has focused on robots entertaining us by achieving various types of performance such as dancing [4], soccer games [3] and so on. And there have been few studies on physical assistance by these robots.

Taking these situations into consideration, we started research on letting a small-sized humanoid robot physically support our daily life. As a first step toward such a goal, we tried to develop a function in which a small-sized humanoid robot would pick up and carry a light object. For this purpose, it is necessary for the robot to achieve the following functions:

(1) To find an object to carry,
(2) To approach the object and stop at the object so that it can pick it up,
(3) To grasp the object,
(4) To carry the object.

In the following, we will describe a system that fulfils these functions.

3 System Construction

3.1 Platform Robot

As a platform robot for this task, we have adopted a commercially available small-sized humanoid robot called "Tai-chi" shown in Fig. 1. Table 1 shows the specifications of the robot. Tai-chi has a total of twelve servo motors for its legs with each leg having six motors, and a total of 8 motors for its arms with each arm having four. Furthermore, Tai-chi has one motor for its head. Therefore, Tai-chi has a total of 21 motors. Tai-chi can do basic movements such as going forward, going backward, turning to the right, and turning to the left. In addition, users can develop various kinds of movements freely by using a special software called "Motion Editor," dedicated to motion creation.

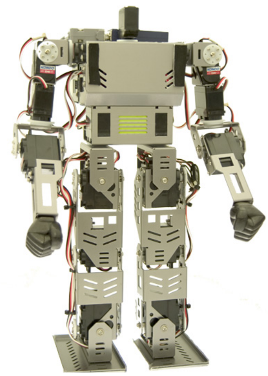


Fig. 1. Humanoid robot used for the research

Table 1. Specifications of humanoid robot

Size/Weight	34 cm / 1.7 kg
Degree of flexibility	22 (12 legs, 8 arms, 1 waist, 1 head)
CPU	SH2/7047F
Motor	KO PDS-2144, FUTABA S3003, FUTABA S3102, FUTABA S3103
Battery	DC6V

3.2 Achievement of the Search Function

As is described in Chapter 2, the system should have the following functions:

(1) To find an object to carry,
(2) To approach the object and stop at the object so that it can pick it up,
(3) To grasp the object,
(4) To carry the object.

As an object to detect and carry, we adopted a small-sized ball with a color of blue, red, or yellow. To give the system the capability of searching/finding objects using image processing, we installed a small CCD camera in the head of Tai-chi. The ball to be detected is decided beforehand as one of the three kinds of balls with red, blue or yellow colors. Then Tai-chi detects the ball utilizing image processing and moves to the location of the ball.

3.2.1 Overview of the Process

First we explain the whole process briefly. Figure 2 illustrates the flow chart of the whole process. At first an image captured by the camera attached to the head of the robot is sent to the PC using a wireless connection. Then the image is transformed into a binary image using the threshold that is established beforehand using the HSV color system. Next, noise is removed from the image using a median filter, and then using labeling processing groups corresponding to a specific color are detected. Next, a group with the largest area is decided as the area corresponding to the ball. Finally, based on the position of the center of the ball image in the image frame, the distance and the direction of the object from the robot is judged. And combining the motion of going forward and turning left/right, the robot approaches the object. In the following, details of the processing will be described.

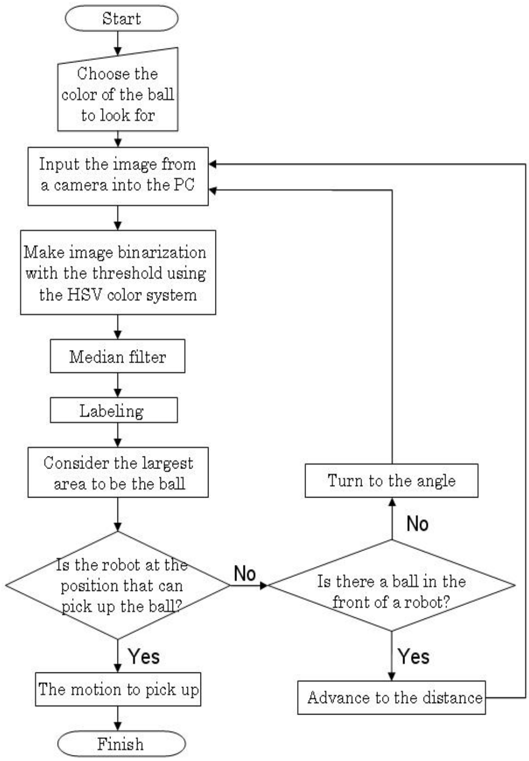


Fig. 2. Flowchart of the search function

3.2.2 A function to Detect a Ball

The first processing in the whole process is performed by detecting the color information assigned beforehand (either red, blue, or yellow) from the original image to find the ball. The original image that has been sent from the camera is based on RGB color information. The detection precision becomes unstable by a change of the illumination where the threshold for detection is based on this RGB information. This is because the RGB information is a color system expressed by a combination of the brightness of each of three primary colors. So the RGB information is not suitable for color detection for a color expressed as the combination of three basic colors. Therefore, we adopted the HSV color system consisting of the combination of H (Hue), S (Saturation), and V (Value) proposed by A. Smith [8]. Changing the lighting condition in many ways, we decided the threshold for the color detection. Using these thresholds, the original image is transformed into a black & white image and the area corresponding to the appointed color is detected from the original image. As noise is still contained in this image, a 3x3 median filter is processed on this image. Then using labeling processing, several areas, one of which is considered to be the ball, are detected and the labeled area with the largest space is detected from the original image as the area corresponding to the ball.

3.2.3 A Function to Move to the Position of the Ball

Then the robot has to recognize the distance and the angle to the ball to move to the position of the ball. Because we use only one camera in this study, we cannot use a method to measure the distance to an object with a stereo camera. Therefore, we fixed an installation position and the installation angle of the camera in this study and developed an algorithm to calculate the distance and direction to an object based on the position of the center of the ball image in the original image.

We set the camera angle so that when the robot was at the position where it could pick up a ball, the image of the ball is at the bottom line of the image. In this case the camera was able to obtain the image about 120 cm away from the position of the robot. In addition, the view of the camera right and left is about 45 degrees. Then measuring a relationship between a ball position in a real space and a a ball position in the image captured by the camera, we developed a map on the image that indicate the robot what to do. Figure 3 illustrates the map. Under the condition mentioned above, we explain the algorithm by which the robot moves in front of a ball. At first the robot performs image processing at its initial position, and the robot performs the turn using the map when the ball is not precisely in front of the robot. It performs the image processing again after it finishes a turn. If a ball is in front of the robot, it advances for the distance to the ball. Although the robot should go straight on precisely with the calculated distance, actually there is a case slipping off to the right or left.

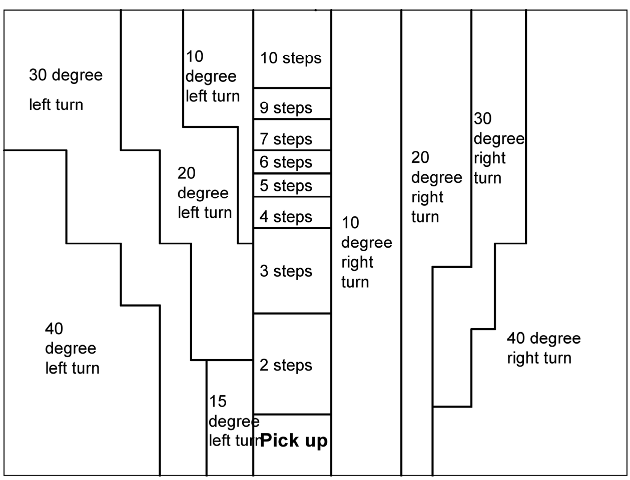


Fig. 3. What to do map on a captured image

Therefore, the robot performs the image processing again after having advanced and it performs an additional turn or progress if necessary. Thus by repeating these movements, the robot moves to the position where it can pick up the ball. Then the robot performs the movement to pick up the ball after having moved.

3.3 Achievement of the Grasping Movement

In this study, we used "motion editor," which is the software dedicated to the humanoid robot "Tai-chi" to make special movements of the robot. The most general and stable movement when we grasp an object and lift it is squatting down, leaning our upper body a bit forward, then grasping the object by stretching our hand forward, then standing up holding the object. We carried out a preparatory experiment to check whether the robot could really pick up a ball with the movement that we developed using the motion editor. Then it was found that the robot was able to pick up the ball when the position of the ball was just in front of the robot or a little to the side of the right position. However, the robot was not able to pick up the ball when the position of the ball was slightly far from the right position. This is because the form of the hand of Tai-chi is a fist and was not able to grasp a ball well. Therefore, we designed another hand form so that it could hold an object stably. Thus the success rate for the robot in picking up and lifting the ball was greatly improved.

3.4 Development of the Walking Motion Holding a Ball

The robot Tai-chi is a compact small-sized robot and has only two modes for achieving motions. One is basic motions such as walking forward, walking backward, turning left, and turning right. These basic motions are achieved based on real-time calculation of motions based on inverse kinematics [5]. On the other hand, other motions dedicated to some special behaviors such as kicking, bowing, and other emotional behaviors are achieved by reading out the data corresponding to the behavior. These data are, as described before, designed and developed using a special software called motion editor. In this study we have to develop a motion by which a robot walks holding a object. Therefore, when a robot walks, the lower part of the body is controlled by a program as before, and it is necessary for the upper part of the body to continue the posture of the last key frame of the motion mentioned in 3.4. To achieve this system, we separated the program of the walking motion into two programs, one of which controls the motion of the upper body and the other the lower body. And we developed a program in which the control of the upper body and the lower body are carried out separately. This new program has an advantage that by combining different motions, new kinds of motions can be developed easily. Logically if there are n motions for upper and lower bodies, n^2 motions can be created using this method. Although we have to be careful not to let the robot fall down, this method has a great advantage when we want to develop complicated motions such as dancing motions. Thus by this method we could achieve the movement of a robot walking with a ball.

4 Evaluation Experiment and Results

4.1 Experiment Results and Consideration

With this system, we tested how correctly the robot could pick up the ball of the color that we appointed beforehand. For the position of the ball, we selected five places, each of which was less that one meter from the robot and within 30 degrees left/right from the front of the robot. For each of these positions we carried out a test for each ball twice. In this study, the robot does not decide all of its behavior patterns at the initial position. Instead, at the place to where the robot moves, it carries out the decision process using image processing and chooses the next action, and by repeating this process it approaches the object. Therefore, the success rates would change depending on the color of the ball and also depending on the location of the ball. Also the time required for approaching the ball and picking up the ball would change depending on these conditions. Therefore, we tried to evaluate these factors.

First, Fig. 4 illustrates the graph of the success rate for each color of the ball. According to Fig. 4, the success rate of the blue ball was the worst, and the success rate of the yellow ball was the best. The one failure trial for the red ball was because the robot lost the location of the ball and could not reach the ball position. On the other hand, in the case of the blue ball, even though it could approach the ball, when it tried to pick up the ball, it made a mistake. The reason for this fault is considered to be as follows. In this case, when the robot came near the ball, because of the reflection of light, only the lower part of the blue ball was extracted using the color information and recognized as a ball. Thus this incorrect image processing led to the incorrect judgment of the ball position and then to the failure to pick up the ball.

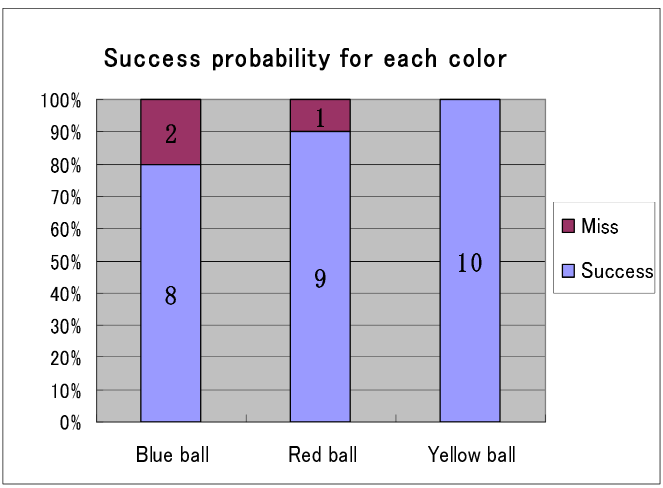


Fig. 4. Success probability for each color

Next, Fig. 5 shows the graph of the success probability for each place where the ball is placed.

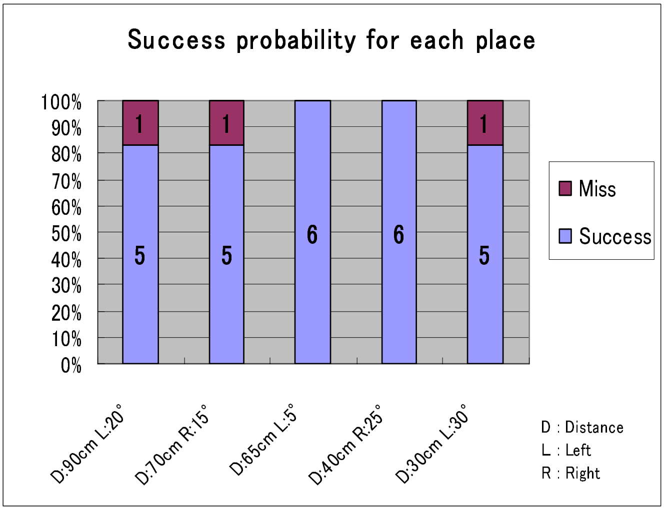


Fig. 5. Success probability for each place

According to Figure 5, there were three failures; in the case of distance 30 cm / left 30 degrees, in the case of distance 70 cm / right 15 degrees, and in the case of distance 90 cm / left 20 degrees. These results do not show big deflection, and it is thought that the success rate would not be affected much by the place condition.

Finally Fig. 6 shows the graph of the average time required until picking up a ball.

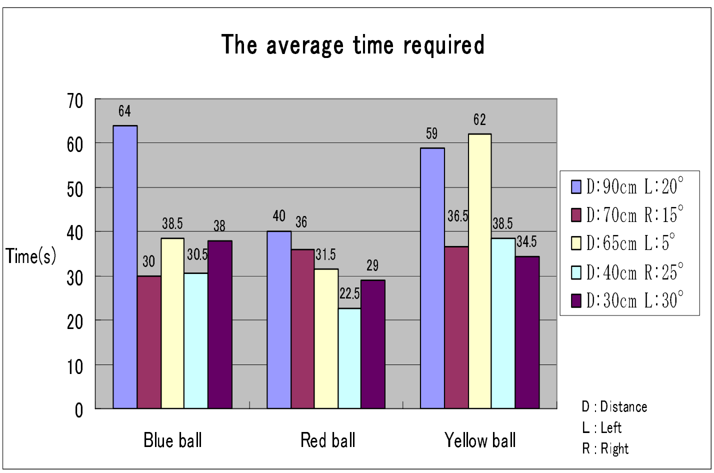


Fig. 6. Average time required

The values associated in the graph show the mean time required for the robot to approach the ball from the initial position and pick up the ball. The three failed trials described above are not counted in this case. As a matter of course, it took long time to arrive at the most distant place (distance 90 cm / left 20 degrees) from the initial position. However, there were not big differences of time between the other four places. The major reason for this would be the following. There is not an appropriate command for the robot to change direction. Instead, the robot tries to change direction by doing stamping. Thus it is difficult for it to adjust to the precise angle by only one trial. This means that the robot could adjust to the precise angle by repeating the direction change movement by stamping several times.

Next let's look at the difference in the required time depending on the color of the ball. It took the longest time for the robot to pick up the yellow ball, and the shortest time to pick up the red ball. The reason for this is that there is a difference in recognition precision depending on the color of the ball. For example, in the case of a yellow ball, there were cases in which it took much time to perform the fine angle adjustment because the extraction of the ball region became unstable because of the light reflection, and only part of the ball was extracted as a ball region. On the other hand, the robot was able to recognize the red ball well regardless of the distance to the ball in contrast with a yellow ball.

We understood that there was a big difference in recognition precision depending on the color of the ball. Especially when the lighting condition changed, the recognition of a ball became difficult.

4.2 Future Study

Even though a fairly simple search algorithm using image recognition based on color information was adopted, a fairly high ball detection rate was obtained. At the same time we found that the recognition precision of the ball changes depending on the lighting condition and the viewpoint. Therefore, it is necessary to improve the image recognition system by introducing object recognition using its own form. Another problem is the fine tuning of the robot movement both for the distance and the angle. The robot has simple commands for the locomotion motions. For going forward and backward, the prepared commands are only to go forward/backward with the appointed steps. Also for angle change, the prepared commands are turn left/right by performing stamping with the appointed repetition. Thus for the locomotion of a precise distance and angle change with a precise angle, it is necessary for the robot to carry out these commands repeatedly, thus requiring a rather long time for the robot to achieve the required action. Coping with this problem is an important research topic. Another issue for shortening time is image processing while walking. As the locomotion of the humanoid robot is achieved by walking using two legs, the image captured by the camera while walking is not stable. Because of this, we adopted an algorithm in which the calculation of the distance/angle based on image processing is carried out when it stops. This repetition of moving then stopping and direction/angle calculation is another reason for the long time required. Therefore, it is necessary to try to develop an algorithm for carrying out image processing while walking by canceling out image shake.

5 Conclusion

In the future it is expected that humanoid robots will be introduced widely into our society. In that case, it is expected that humanoid robots would support our life physically in addition to entertaining us. As a first step toward this goal, in this study we aimed to create a small-sized robot that would support us physically by carrying such objects as coffee cups or newspapers. Based on this basic concept, we tried to achieve functions by which a small-sized humanoid robot discovers an object, approaches it, picks it up, and carries it. Then we carried out experiments to evaluate the capability of the functions achieved by the small-sized robot. The experimental results showed the robot could pick up a ball with considerable precision and was able to achieve the aim of finding an object and carrying it.

On the other hand, we discovered some points that needed improvement while we pushed forward with the study. Primarily, the robot is considerably uneven in recognition precision in indoor environments. In addition, it is necessary to improve the robot on the hardware side. High functionality of the robot is necessary, so we want to examine these points and will continue to study them in future research.

References

1. AIBO Official Site, `http://www.jp.aibo.com/`
2. ASIMO OFFICIAL SITE, `http://www.honda.co.jp/ASIMO/`
3. Friedmann, M., Kiener, J., Petters, S., Thomas, D., von Stryk, O., Sakamoto, H.: Versatile, high-quality motions and behavior control of humanoid soccer robots. In: Proceedings of Workshop on Humanoid Soccer Robots of the 2006 IEEE International Conference on Humanoid Robots, pp. 9–16 (2006)
4. Shinozaki, K., Iwatani, A., Nakatsu, R.: Concept and Construction of a Robot Dance System. The International Journal of Virtual Reality 6(3), 29–34 (2007)
5. Wama, T., Higuchi, M., Sakamoto, H., Nakatsu, R.: Realization of Tai-chi Motion Using a Humanoid Robot. In: Entertainment Computing. LNCS, pp. 14–19. Springer, Heidelberg (2004)
6. Okada, K., Ogura, T., Haneda, A., Kousaka, D., Nakai, H., Inaba, M., Inoue, H.: Integrated System Software for HRP2 Humanoid. In: Proc. of International Conference on Robotics and Automation (ICRA 2004), pp. 3207–3212 (2004)
7. Pineau, J., et al.: Towards robotic assistants in nursing homes: Challenges and results. Robotics and Autonomous Systems 42(3-4), 271–281 (2003)
8. Smith, A.R.: Color gamut transformation pairs. Computer Graphics 12, 12–19 (1978)

Interactive Multimedia Contents in the IllusionHole

Tokuo Yamaguchi, Kazuhiro Asai,
Yoshifumi Kitamura, and Fumio Kishino

Graduate School of Information Science and Technology, Osaka University,
2-1 Yamada-oka, Suita, Osaka 565-0871, Japan
{yamaguchi.tokuo,asai,kitamura,kishino}@ist.osaka-u.ac.jp

Abstract. This paper proposes a system of interactive multimedia contents that allows multiple users to participate in a face-to-face manner and share the same time and space. It provides an interactive environment where multiple users can see and manipulate stereoscopic animation with individual sound. Two application examples are implemented; one is location-based content design and the other is user-based content design. Both effectively use a unique feature of the IllusionHole, i.e., a location-sensitive display device that provides a stereoscopic image with multiple users around the table.

Keywords: 3D user interface, entertainment computing, tabletop display, interactive, CSCW, game, stereoscopic display.

1 Introduction

Interactive multimedia contents combined with images and sounds have become widespread in various fields. These contents allow for abundant expression by users. For example, using virtual reality technologies, we can enjoy rich high-immersion experiences using multimodal interactions with various information channels such as visual, auditory, or tactile perception. Moreover, we can easily communicate at remote locations through networks and simultaneously share images and sounds in a huge virtual environment. There are, however, some limitations; for example, each of multiple users will not be able to get such rich high-immersion experiences simultaneously because communication channels through the network are not sufficient to convey non-verbal information that is important for natural communications.

In order to achieve natural interactions or collaborations for multiple users, tabletop approaches have been focused on. These approaches can provide a common workspace where multiple users may interact with each other and maintain awareness of what the others are doing. Based on this idea, many researchers explore the role of novel interfaces and interaction paradigms in the context of applications for entertainment. There is, however, little reported in the literature about tabletop displays that allow multiple users to enjoy stereoscopic animation with individual sound.

In this paper, we propose a multimedia content display system for co-located multiple users with which they can enjoy interactive multimedia content (such as a game) or work cooperatively through natural face-to-face communications while sharing the same time and space. The system is based on the IllusionHole [7]. It provides an

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 116–121, 2008.

interactive environment where multiple users can see and manipulate stereoscopic animation with individual sound. The content is basically common to all users; however, it is slightly varied or personalized, such as direction of animations or volume of sounds, according to the user's interactions or dynamically changing positional relationships with other users. Two types of application examples are implemented. Both provide individual animation and sound with a corresponding user, as well as the location-based content design and the user-based content design.

2 Related Work

Recently, some studies have shown that the combination of visual and auditory cues enhances the sense of immersion in virtual reality or interactive entertainment applications. This section outlines a variety of interactive stereoscopic displays and interfaces using auditory feedback.

To view objects from different angles by moving one's head provides humans with important clues for spatial cognition. One of the most reasonable ways to create a multi-user interactive stereoscopic display is to install a horizontal screen in a table [1]. This is the most effective way to view stereoscopic images from the vantage point of individuals standing around a table. And to use optical equipment that includes a parallax barrier [4], mirror [2, 11] and revolving screen [5] allows multiple users to observe the stereoscopic images with motion parallax in any direction. In addition, multiple users are able to directly point to a particular part of the stereoscopic images in the IllusionHole [7].

Many co-located collaborative applications often present information to users through auditory channels as well as through visual feedback. For providing awareness of the users' actions, large-screen tiled displays use auditory information when users perform gestures or move objects [10]. Morris et al. found that increasing collaboration can result when individual, rather than public, auditory feedback is provided [9]. For music tabletops, the reacTable uses physical objects to represent parts of a modular synthesizer [6], and the Audiopad is a composition and performance instrument for electronic music that tracks the positions of objects on a tabletop [12]. Jam-O-Drum allows users to collaboratively create music around a circular tabletop [3], and with Multi-Audible [8], multiple users have a portable device to hear different audio information during the interaction.

3 System Configuration

In this section, we detail the configuration of an interactive multimedia content system that allows multiple users to see and manipulate stereoscopic animation with individual sound.

3.1 Overview

Our proposed multimedia content system allows multiple users standing around a table to interact with position-specific stereoscopic animation with individual sound. The content itself is basically common to all users; however, it is slightly varied or personalized according to users' interactions. We have implemented two different types of content design frameworks; i.e., location-based content design and user-based content design. In the former framework, stereoscopic animation and sound changes interactively according to the location where a user stands. The environment around the table is divided into an adequate number of domains. If a user comes into one domain from another, the content that is currently displayed to the user changes according to the domain. On the other hand, in the user-based content design, the multimedia content changes according to the user's positional relationships with other users. In this case, each user is also provided with interactive animation and sound, and they are changed when the surrounding environment changes, such as when another user comes closer or moves away.

3.2 Implementation

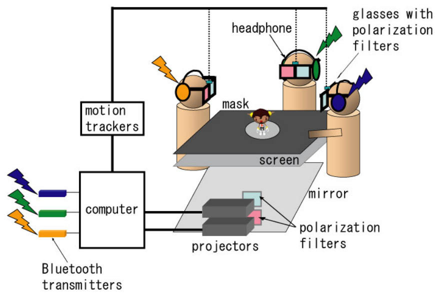


Fig. 1. System configuration

A prototype system was established using the IllusionHole with polarization filters (see [7] for more details). Based on the IllusionHole, the prototype system allows multiple users to see interactive stereoscopic animations with adequate motion parallax. In addition, Bluetooth-enabled wireless headphones are installed to provide individual sound for corresponding users. The system configuration is shown in Figure 1. The viewing position of each user is detected using a 3D tracker device (IS-600 Mark 2 ultrasonic beacon, made by Intersense) and stereoscopic images are displayed using parallax images for both eyes, which are calculated for corresponding display regions. The display regions of multiple users may overlap each other if the number of users increases and neighboring users stand too close to one another. So, the prototype system is designed for three users. Each user wears a pair of circularly polarized glasses and a wireless headphone corresponding to an output sound channel.

A direct graphics library is used to manage and show stereoscopic animations of 3D characters. In addition, a direct sound library is used for generating multi-channel sound for individual sound. Figure 2 shows the output flow of individual sound. The direct sound library supports multiple buffers of multi-channel sound, and also supports the mixing of multiple sound buffers to play these sounds. Multiple buffers are assigned to each individual user, and these buffers have some slots for writing sound data which output from each channel (ch1 to ch8). The first slot "H" includes a header for the buffer. The sound data for only a particular user is written at the output channel assigned to her/him (e.g., user A is assigned ch1 and ch4, and user C is assigned ch3 and ch6), and null data is written in the remaining channel for silent output. By mixing the buffers in which sound data is written for each user, and controlling sound volume and the starting or stopping of sound, we manage the individual sounds.

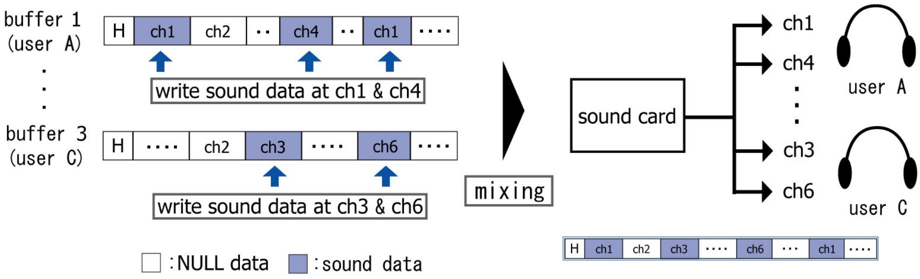


Fig. 2. Output flow of individual sound

4 Application Examples

We introduce two application examples that allow multiple users to communicate interactively in the same place at the same time using this configuration. "Onstage Demo of the IllusionHole" is an example of the location-based content design, and "Baa Baa White Sheep" is an example of the user-based content design.

4.1 Onstage Demo of the IllusionHole

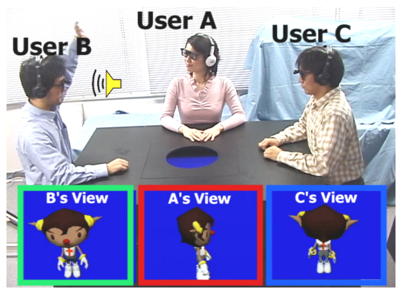


Fig. 3. Snapshot of Onstage Demo of the IllusionHole experienced by three users

In this application, the character shown at the center of the IllusionHole talks about functions and features of the IllusionHole itself. Figure 3 shows a snapshot of this application experienced by three users and different stereoscopic animations observed by individual users standing to the left, center, and right of the IllusionHole. The particular moment of Figure 3 shows that the 3D character faces to the left; therefore, only the left user (user B) hears the individual announcement as "Raise your hand" by the character; however, the other users (users A and C) cannot hear anything. In this application example, the character turns to other directions and talks about the IllusionHole similarly. The IllusionHole is a location-sensitive display device that provides a stereoscopic image with multiple users around the table, and each of the users can observe the same virtual object from a different direction. Therefore, this application example effectively uses this feature of the IllusionHole.

4.2 Baa Baa White Sheep

This application is designed so that a character in the IllusionHole corresponds to a user's actual position and movements in the physical world. A snapshot of users enjoying this application is shown in Figure 4. By simply moving around the display, each user can manipulate his/her own character without using devices such as game

Fig. 4. Snapshot of users enjoying Baa Baa White Sheep

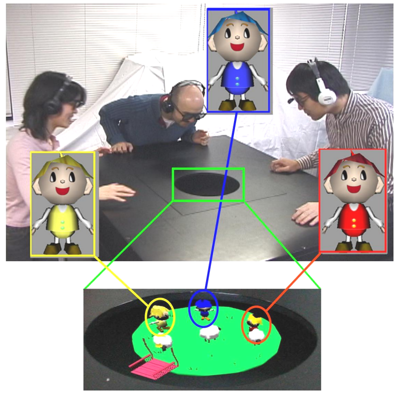


Fig. 5. Correspondence between users and characters

controllers or mice. In the scenario of this application, users have to cooperatively drive sheep into the fold taking into consideration the relative positions of other users, while the sheep tries to escape from the users. Figure 5 shows the correspondence between users and characters. If a user starts moving, the system detects his/her motion and changes the animation of the corresponding character to "walking" from "standing." At the same time, the user can hear the sound of footsteps fitting the "walking" animations. Moreover, the other users can also hear the sounds of footsteps according to the relative position with the user. And if a user comes closer to the sheep, only he/she hears the sound of the bleat. In this way, users enjoy interactions with a virtual world using physical movements and the relative positions to others, while the user feels the character as another human being. The video figure shows details of this application.

5 Conclusions

In this paper, we proposed a system for multiple co-located users to enjoy interactive multimedia contents while sharing the same time and space. We described the design approach and implementation of the system followed by two application examples. In the future, we are planning to explore new content suitable for stereoscopic animation with individual sound using physical body gestures or utterances as well as using physical movements and relative positions to others. Moreover, we also plan to look into studying how users perform and cooperate with a variety of personalities and leadership qualities in these environments.

Acknowledgments

This study was supported in part by "Global COE (Centers of Excellence) Program" of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

1. Agrawala, M., Beers, A.C., Frohlich, B., Hanrahan, P.: The two-user responsive workbench: support for collaboration through individual views of a shared space. In: Proc. of SIGGRAPH, pp. 327–332 (1997)
2. Bimber, O., Frohlich, B., Schmalstieg, D., Encarnacao, L.M.: The virtual showcase. IEEE Computer Graphics and Applications 21(6), 48–55 (2001)
3. Blaine, T., Perkis, T.: The Jam-O-Drum interactive music system: a study in interaction design. In: Proc. of the 3rd Conference on Designing Interactive Systems, pp. 165–173 (2000)
4. Endo, T., Kajiki, Y., Honda, T., Sato, M.: Cylindrical 3D display observable from all directions. In: Proc. of the 8th Pacific Conference on Computer Graphics and Applications, pp. 300–306 (2000)
5. Favalora, G., Dorval, R.K., Hall, D.M., Giovinco, M., Napoli, J.: Volumetric three-dimensional display system with rasterization hardware. In: Proc. of SPIE, vol. 4297, pp. 227–235 (2001)
6. Jordà, S., Geiger, G., Alonso, A., Kaltenbrunner, M.: The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In: Proc. of the 1st International Conference on Tangible and Embedded Interaction, pp. 139–146 (2007)
7. Kitamura, Y., Nakayama, T., Nakashima, T., Yamamoto, S.: The IllusionHole with polarization filters. In: Proc. of ACM Symposium on Virtual Reality Software and Technology, pp. 244–251 (2006)
8. Kusunoki, F., Eguchi Yairi, I., Nishimura, T.: Multi-Audible table for collaborative work. In: Proc. of ACM SIGCHI International Conference on Advances in computer entertainment technology, pp. 67–73 (2004)
9. Morris, M.R., Morris, D., Winograd, T.: Individual audio channels with single display groupware: effects on communication and task strategy. In: Proc. of ACM Conference on Computer Supported Cooperative Work, pp. 242–251 (2004)
10. Müller-Tomfelde, C., Steiner, S.: Audio-enhanced collaboration at an interactive electronic whiteboard. In: Proc. of the 2001 International Conference on Auditory Display, pp. 267–271 (2001)
11. Otsuka, R., Hoshino, T., Horry, Y.: Transpost: all-around display system for 3D solid image. In: Proc. of ACM Symposium on Virtual Reality Software and Technology, pp. 187–194 (2004)
12. Patten, J., Recht, B., Ishii, H.: Audiopad: a tag-based interface for musical performance. In: Proc. of Conference on New Interface for Musical Expression, pp. 24–26 (2002)

Creating an Emotionally Adaptive Game

Tim Tijs[1], Dirk Brokken[2], and Wijnand IJsselsteijn[3]

[1] User-System Interaction Program, Department of Industrial Design
Eindhoven University of Technology, P.O.Box 513, 5600MB, Eindhoven, The Netherlands
[2] Philips Research Laboratories Europe
High Tech Campus 34, 5656 AE, Eindhoven, The Netherlands
[3] Human-Technology Interaction Group, Department of Technology Management
Eindhoven University of Technology, P.O.Box 513, 5600MB, Eindhoven, The Netherlands
t.j.w.tijs@tue.nl, dirk.brokken@philips.com,
w.a.ijsselsteijn@tue.nl

Abstract. To optimize a player's experience, an emotionally adaptive game continuously adapts its mechanics to the player's emotional state, measured in terms of emotion-data. This paper presents the first of two studies that aim to realize an emotionally adaptive game. It investigates the relations between game mechanics, a player's emotional state and his/her emotion-data. In an experiment, one game mechanic (speed) was manipulated. Emotional state was self-reported in terms of valence, arousal and boredom-frustration-enjoyment. In addition, a number of (mainly physiology-based) emotion-data features were measured. Correlations were found between the valence/arousal reports and the emotion-data features. In addition, seven emotion-data features were found to distinguish between a boring, frustrating and enjoying game mode. Taken together, these features convey sufficient data to create a first version of an emotionally adaptive game.

Keywords: Adaptivity, personalization, computer games, affective loop, psychophysiology, emotions.

1 Introduction

For creating entertaining computer games[1], gameplay is considered to be of key importance ([1], [2]). In absence of a broadly accepted definition of gameplay, we focus here on one frequently mentioned element of it, which is challenge. The process of optimizing a game's challenge is referred to as *game balancing* or *difficulty scaling*. That is, changing parameters in order to avoid that the player gets frustrated because the game is too hard, or gets bored because the game is too easy [3]. In this study, we have investigated the relations between a game's difficulty level and the interplay between the emotions boredom, frustration and enjoyment (Fig. 1-left panel). These

[1] From this point onward referred to as "games". In the present research, the term includes electronic games played on personal computers as well as those played on modern dedicated video game consoles, such as Sony's Playstation 3, Microsoft's Xbox 360 and Nintendo's Wii.

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 122–133, 2008.

relations strongly differ per individual, for example influenced by a player's skill level. For instance, a difficulty level that is found enjoyable by a novice might be boring for expert players. Games therefore need psychological customization techniques [4], such as difficulty adaptation, to optimize the experience. Since many years, game designers aim to provide some customization, for example by letting players choose a difficulty level upfront or including progressive difficulty levels during gameplay, based on a player's performance.

More advanced methods that work in real-time are less common. One difficulty adaptation mechanism, frequently applied in racing games, is called rubber banding [5]: When falling behind, the player suddenly gets an enormous boost in speed, which allows for catching up again (and vice versa for the competing cars). However, game adaptation that is solely based on in-game behavior can only have limited success, because there are many different types of players [6]. Each type of player has his/her own goals, preferences and emotional responses when playing a game. Hence, for optimizing the players' experiences, successful psychological customization requires a game to take the emotional state of the player into account. Games should become emotionally adaptive (Fig. 1– right panel).

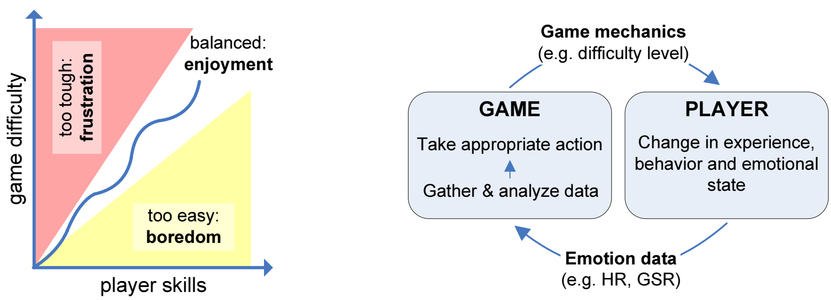


Fig. 1. Left panel: Game balancing (Adapted from [1]). Right panel: The emotionally adaptive games loop, inspired on the affective loop [7].

1.1 Emotionally Adaptive Games

The importance of emotions in computing is widely argued for (e.g. [8]). Emotion theorists differ over a discrete versus a dimensional model. The "discretionists" (e.g. [9]) argue for basic discrete emotions, such as anger, fear, sadness and happiness, as unique experiential states. The "dimensionalists" (e.g. [10]), on the other hand, look at emotions in terms of a two-dimensional space consisting of *valence* ("pleasantness") and *arousal* ("activation"). Sometimes *dominance* is added as a third dimension.

Effective human-computer interaction from an emotions perspective works in terms of an "affective loop" [7]. A similar feedback loop in a games context is described by [11]. Inspired by their work, Fig. 1 (right panel) shows a schematic view on the functioning of an emotionally adaptive game. By providing the right game mechanics [12] (e.g. audiovisuals, narrative, challenge), the game influences the player's experience, behavior and emotional state. Ideally, during play, the emotional state of the player (measured in terms of *emotion-data*), is continuously being fed back to the game so that the game can adapt its mechanics (e.g. difficulty level) accordingly in real-time. This all is done to create the optimal experience (which is referred to in literature as e.g. *flow* [13] or *immersion* [14]). Previous research attempts to create emotionally adaptive software have mainly focused on tutoring systems and workload / performance optimization (see e.g. [15]). Fewer attempts have been made to incorporate a closed-loop mechanism in a games context. Takahashi et al. [16] and Rani et al. [17] created a game that was found to improve player performance by adapting difficulty level to player's physiological state. Concept validation claims of these both studies were, however, based on a limited number of participants. Besides these attempts, a number of *biofeedback games* have recently been developed, which have some integration of a player's physiological data into the game (e.g. [18], [19] and [20]). These games however focus on stress manipulation rather than optimization of gameplay experience. Probably closest to the present project's scope is the work of Saari and colleagues, who created the Mind-Based Technology framework for psychological customization [21]. They have further elaborated this in a games context (e.g. [4], [22]) and are currently developing an emotionally adaptive game demo.

As a first step in creating an emotionally adaptive game, system input and output need to be specified in further detail. Regarding output (emotion-data), Saari et al. [22] provide an extensive discussion of possible elements to be adapted, structured by "psychologically validated templates". We have adopted a rather straightforward and intuitive "template": *Game speed*. We will manipulate the game's speed to influence the player's emotional state (the interplay between boredom, frustration and enjoyment, Fig. 1-left panel). Regarding system input (emotion-data), Öhman [23] distinguished three categories of emotion measures: Self-reports, overt behavior and physiological responses. Self-reports are frequently used for assessing players' emotions and experiences [5] but not suitable (since too obtrusive) for real-time application in a game. Regarding overt behavior, potentially useful techniques for measuring boredom, frustration and enjoyment are facial emotion tracking [24] and the analysis of posture and pressure exerted on the game controls [25]. Regarding physiological responses, there is an extensive field with many research findings in psychophysiology. Although the research is done in varying contexts with sometimes contradicting results, it is considered a highly interesting field for analyzing emotions in games. We have limited ourselves to the methods described below.

Regarding cardiovascular (heart) activity, tonic (long-term, as opposed to phasic) heart rate (HR) is known to increase with sympathetic nervous system activity, such as emotional arousal and cognitive effort and stress. On the other hand, increases in attention (mediated in the parasympathetic nervous system) lead to a decreased heart rate [26]. [27] found HR features to correlate with self-reported fun in games. Heart rate variability (HRV) is considered an index for mental effort (e.g. [28]). Some researchers (e.g. [29]) consider the percentage power in the low-frequency (LF) 0.07-0.14 Hz range as a particularly effective index for task-related mental effort / sympathetic activity. Respiratory responses are analyzed to control for respiratory artifacts in e.g. HRV (a phenomenon known as respiratory sinus arrhythmia). Respiration may, however, also be used as a measure itself, e.g. for investigating stress and mental load [30]. Electrodermal activity (EDA) concerns the electrical resistance of the skin, also known as Skin Conductance (SCL, SCR) or Galvanic Skin Response (GSR). Skin conductance level is known to increase with information processing and the frequency of non-specific skin responses increases with arousal [26]. Electromyography (EMG) is a technique for measuring muscle activity; electric potential is being generated when muscle cells contract. Facial EMG is frequently used as a metric for valence. The most frequently analyzed facial muscles in this context are the orbicularis oculi (OO, used for closing the eyelids), zygomaticus major (ZYG, smiling) and corrugator supercilii (CORR, frowning). Most studies find positive correlations between valence and the OO and ZYG muscles, and a negative correlation between valence and CORR muscle (see e.g. [31], [32], [33]). In addition to the above findings, there are also a considerable number of studies without significant findings [34]. Because of the large differences in physiological responses between individuals and within individuals over time (autonomic response stereotypy principle, see e.g. [15], [35]), some researchers (e.g. [36]) argue for normalizing physiological data to facilitate a group analysis of the data. Additionally, affective systems should employ a battery of physiological features for accurate emotion predictions (e.g. [37]), and should allow for user-control for the sake of autonomy, privacy and interpretation of the data [41].

Because of the context-dependency of physiological responses, a two-stage approach was adopted. The purpose of the current initial study is to investigate physiological and other affect-related responses in relation to an experimentally induced change in game mechanics. Note that in this study the affective loop is not yet closed, that is, real-time affective indicators are not yet directly influencing the game mechanics. This will be the purpose of phase 2 of our work. The research question for the current invesigation evolved around the components of our framework (Fig. 1-right panel): *What game mechanics (speed settings) lead to what kind of emotional state, and what emotion-data is this accompanied by?* This was investigated by means of a controlled experiment, as explained in the next section.

2 Method

2.1 Participants

In total 24 adults aged 23-46 (mean age 29.7) participated in the experiment; 19 males and 5 females. The participants played computer games (console, pc, arcade, and internet) for an average of 12.5 hours per month. All had played Pacman before. The participants were rewarded a €5 voucher for participation plus another €5 voucher for the final top-5 scores. All were informed in advance about this incentive. A within-subjects design was applied using the game's speed-mode (3 levels: slow, fast and normal) as an independent variable. The order in which participants were confronted with the two non-normal speed modes (i.e. slow and fast) was counterbalanced over participants. A dual-PC setup was used to separate the game and questionnaire from the measurement software, to ensure the game and computer questionnaire to run smoothly.

2.2 Design, Materials and Procedure

The stimulus (game) was an adapted version of the PC-game Pacman [38]. Pacman is a relatively uncomplicated game without major changes in e.g. audiovisuals during play, which could lead to emotional bias. Moreover, it is a well-known game and is easy to pick up (requiring relatively short practice to minimize learning effects). Pacman also has a rather continuous flow of action (useful when comparing blocks of time) and has been used before in affective computing studies (e.g. [39]). To suit the game to the experiment, some adaptations were made: I) the player stayed in the same level during the experiment, II) objects that were eaten, such as points and pills, returned after a while), III) the speed level changed at preset times (unknown to the player), IV) eating objects increased the player's score, but being eaten meant a strong decrease in score, and V) the overall objective was to score as many points as possible. The choice for manipulating the difficulty parameter speed, instead of e.g. the number of opponents (i.e. ghosts) is based on the fact that the number of normal ghosts constantly changes during gameplay, because of Pacman eating star-shaped pills. Pacman was played using the arrow keys on the keyboard. All participants were instructed to play with their preferred hand.

The dependent variables comprised the player's emotional state and several emotion-data features. To gather demographics and gaming experience (i.e. gaming hours per month), a paper-questionnaire was filled out before the start of the game. Emotional state was measured through self-reporting in a computer-questionnaire (in between moments of play) and a post-game interview. The computer-questionnaire comprised two similar parts of three items, that focused on the minute of play before switching to this questionnaire. The first two items addressed valence and arousal, using Self-Assessment Manikins ([10], adapted to a five-point scale). The third was a multiple-choice item that a subject could use to indicate that (s)he was either enjoyed, bored or frustrated (or that none of these descriptions was applicable). The semi-structured interview focused on the subject's overall opinion of the experiment and his or her preferred speed mode (fast, slow or normal). The emotion-data features that were collected during the game are displayed in Table 1.

Table 1. Emotion-data features analyzed in the present study

Measurement (sample freq in Hz)	Feature	Abbreviation
Blood Volume Pulse (128)	BVP level	BVP
	BVP amplitude	BVP_AMP
	Heart rate	HR
	HRV-LF% (0.04-0.15Hz)	HRV_LF
Respiration (32)	Respiration amplitude	RSP
	Respiration rate	RSP_RATE
	RSP-HR coherence	RSP_HR_COH
Skin Conductance (32)	Skin conductance level	SCL
	Number of SC responses	SCR
Facial Electromyography (1024)	Activity of the CORR	CORR
	Mean amplitude of the CORR	CORR_AMP
	Activity of the ZYG	ZYG
	Mean amplitude of the ZYG	ZYG_AMP
Keyboard Pressure (100)	Keypress-pressure	KEYB

The physiological measurements were done using a wireless NeXus-10 kit. At the PC, the physiological data were stored using Biotrace+ v.1.16[2]. A BVP photoplethysmograph was placed on the ring finger of the subject's non-playing hand. A respiration belt was strapped around the thorax; a maximum of 2 layers of cloth was in between. Skin conductance sensors were placed on the tips of the index and middle finger of the non-playing hand. The Facial EMG electrodes (AgCl, 15mm. diameter) were placed following instructions obtained from [40]. Keyboard pressure was measured by a pressure sensor placed under one of the keyboard's feet (the one closest to the arrow keys); keypresses were identified using a custom-made application. To synchronize the game events with the measurement data, a 3-window capture was implemented on the analysis PC.

Table 2 provides an overview of the test procedure.

Table 2. Test procedure

Activity		**min.**
Welcome, consent form, paper-questionnaire		5-10
Installation of phys. equipm., game instructions, short practice		5-10
Game part 1	Normal speed	2
	Non-normal speed	2
Computer-questionnaire part 1		~1
Game part 2	Normal speed	2
	Non-normal speed	2
Computer-questionnaire part 2		~1
Game part 3	Normal speed	1
Semi-structured post-game interview, debriefing		5
Total		**~30**

2.3 Analysis

The feature analysis process consisted of a number of steps, as explained in Fig. 2.

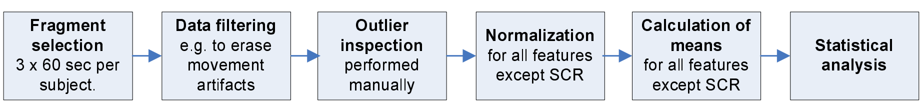


Fig. 2. Analysis process

The three selected fragments per subject are from this point onward referred to as *slow-mode*, *fast-mode* and *normal-mode*. The physiological features were extracted and filtered using the default Biotrace+ detection and filtering algorithms. For the GSR, EMG and KEYB features, additional algorithms were used. The EMG signals (pre-amplified in the Nexus kit) were linearly rectified for analysis, high-pass filtered (Chebyshev Type 2, high-pass, cut-off frequency at 20 Hz) and integrated (smoothened) using a 200 ms running average. For identifying skin conductance responses (SCR) and keypresses from the pressure recordings (KEYB), peak detection

[2] Both NeXus and Biotrace+ are manufactured by Mind Media b.v., The Netherlands.

algorithms developed at Philips Research were used. Normalization of the physiological features, done according to [36], is required for a group analysis, because of the strong inter-subject differences in common physiological responses. These differences were also found for SCR; therefore SCR also requires a form of normalization. However, this could not be accomplished with the formula explained in [36], since one subject had in all three conditions (slow, fast, normal) the same number of identified skin conductance reponses. Therefore it was decided to normalize the SCR values for each subject over the subject's total signal span (3 conditions, thus 3 SCR values) in terms of inverted ranking[3]. Finally, statistical analysis was performed with SPSS v.16.0 software. As already mentioned in chapter 1, the analysis was focused on the relations between the various speed settings, the player's emotional state, and the accompanying emotion-data.

3 Results

Table 3 presents the self-report results from the questionnaire as well as from the post-game interview. Based on the results in Table 3, we draw the following conclusions: I) the participants felt bored in the slow-mode, II) in the fast-mode some were frustrated, others enjoyed themselves, and III) the normal-mode was preferred over the other two speed modes. Speculatively, we conclude the following: the slow-mode was too slow, the fast-mode was for some a bit too fast, but for others the right speed level. The speed level in the normal-mode might not be optimal either, but the players' experiences are better in that mode than in the other two.

Table 3. Self-report results

Computer-questionnaire ("I felt..." - count)				
	bored	frustrated	enjoyed	not applicable
slow-mode	20	2	2	0
fast-mode	2	8	14	0
Post-game interview (preferred speed mode – count)				
slow-mode	fast-mode		normal-mode	
0	1		23	

There was a tendency toward neutral scoring on the SAM (valence and arousal) topics. For instance, the 22 non-enjoyment scores (i.e. reports of boredom or frustration) in the slow-mode were accompanied by only 5 negative valence reports. To investigate the relations between the SAM (valence and arousal) scores and the emotion-data features, both valence scores and both arousal scores for a subject were averaged. A Pearson correlation test revealed two statistically significant results ($p<0.050$), as displayed in Fig. 3. Firstly, arousal correlated with the emotion-data feature BVP ($r=0.449$, $p=0.032$, $N=23$). Since the arousal score of 1 represents the

[3] I.e.: the highest of the 3 SCR values for a subject was awarded the rank 3 and the lowest score of the three scores the rank 1. For instance, the SCR values 3 (slow-mode) ,4 (fast-mode) and 1 (normal-mode) for a subject were respectively transformed into the normalized values 2, 3 and 1. Similarly, a subject's SCR values of 10, 10 and 10 were transformed into 2, 2 and 2.

highest arousal report, this correlation should be interpreted as "*the higher the arousal report, the higher the BVP value of a player in the non-normal game modes*". Secondly, valence appeared to correlate negatively with the ZYG_AMP feature (r=-0,446, p=0,033, N=23). This result should be read as *"the more positive valence reports in the non-normal speed modes were accompanied by larger movements in the zygomaticus major muscle*". When observing this correlation in Fig. 3 (right panel), it seems that two data points (squares in both lower corners) have a strong influence on the correlation effect. Although these data points lay rather far out, they were kept in the analysis because I) both represent an average of two conditions, II) the video recordings of the related participants do not reveal any anomalies and III) strong inter-subject differences in physiological responses are common (e.g. [15]).

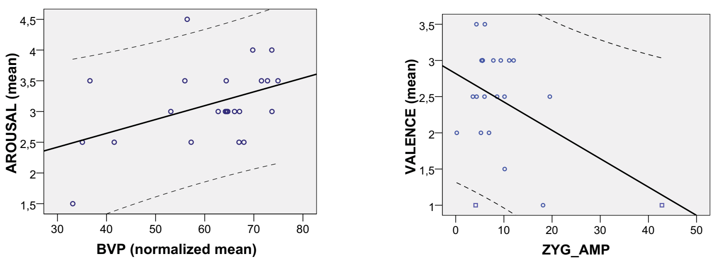


Fig. 3. Significant correlations between the SAM scales and emotion-data features. For the valence and arousal axes: 1=Most positive/highest, 5=least positive/lowest. Dotted lines represent confidence intervals (95%).

Table 4. Emotion-data features that distinguish between the game's speed modes

Feature	Conditions	t	df	p	mode identified (desired action)
SCL	slow-fast	-2.511	22	0.020	slow (speed up)
	slow-normal	-4.197	22	0.000	
SCR	slow-fast	-4.851	22	0.000	slow (speed up)
	slow-normal	-4.343	22	0.000	
HR	slow-fast	-5.177	22	0.000	slow (speed up)
	slow-normal	-5.196	22	0.000	
RSP_RATE	slow-fast	-3.806	22	0.001	slow (speed up)
	slow-normal	-2.388	22	0.026	
CORR_AMP	slow-fast	3.078	22	0.006	fast (slow down)
	fast-normal	-3.592	22	0.002	
ZYG_AMP	slow-fast	-2.673	22	0.014	fast (slow down)
	fast-normal	2.767	22	0.011	
KEYB	slow-fast	-5.136	22	0.000	slow (speed up) & fast (slow
	slow-normal	-3.637	22	0.001	down) & normal (none)
	fast-normal	3.282	22	0.003	

To analyze the effects of speed on the players' emotion-data, a repeated measures ANOVA was performed with speed (3 levels: slow, fast, normal) as a within subjects factor. This provided an overview with a number of significant effects. To analyze these effects (i.e. the differences in emotion-data between the 3 speed modes) in more detail, a series of paired sample t-tests was performed. Significant results ($p<0.050$) are provided in Table 4. For example, Table 4 and Fig. 4 show that the average skin conductance level (SCL) during the slow-mode was significantly lower than during the other two speed modes. Since the participants reported boredom during the slow-mode, our conclusion is that this mode was too slow. In other words, the game should speed up. According to Table 4, keyboard pressure (KEYB) can even help to distinguish between the game's 3 speed levels.

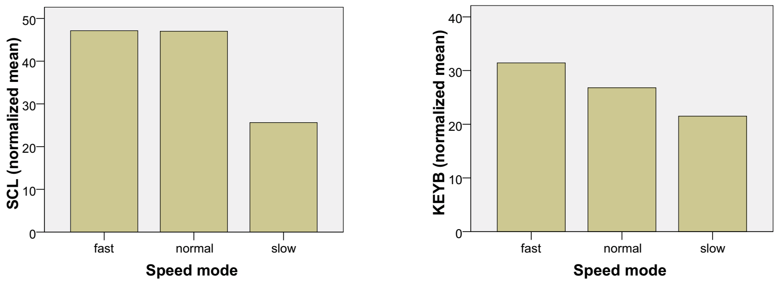


Fig. 4. Significantly different emotion-data SCL (left panel) and KEYB (right panel)

4 Discussion and Further Research

When comparing our findings to other studies that investigate in the relations between valence, arousal and physiological responses, one result seems to contradict most other work: The positive correlation between self-reported arousal and BVP (Fig. 3-left panel). In addition to the central tendency in scoring on valence and arousal, no less than 8 of the 24 participants asked for the exact meaning of the arousal manikins when first confronted with them. Two even thought that the highest-arousal manikin was supposed to depict hunger. In turn, the experimenter described it to them as their level of activation.

The induction of boredom, frustration and enjoyment by manipulation of the game mechanic "speed" was partly successful. Nearly all players indicated to be bored during the slow-mode, but the fast-mode was found more enjoyable than frustrating. There are several possible causes for this. Firstly, the players knew the game speed was going to change. Besides, the speed changes were rather abrupt, and the players knew it only lasted for a limited amount of time. Nonetheless, nearly all participants (see Table 3) considered the normal-mode the most enjoyable of the three speed modes.

As a stimulus, Pacman appeared to be a successful game choice. No major problems (such as crashing software or a stalemate situation in the game) were encountered during the experiment. Besides, from the post-game interviews, we conclude that all participants enjoyed the game from an overall perspective. It was also relatively easy to implement the speed manipulations and customized ruleset in the game.

In the next phase of this project, we will aim to develop an emotionally adaptive version of the Pacman game. This game will use the data collected in the present study as a starting (or: calibration) point for adaptation. For instance, when the emotionally adaptive Pacman game receives emotion-data that matches the data pattern found for the "boring" slow-mode (e.g. relatively low SCL and SCR values, see Table 4), it will either increase its speed or, to preserve player autonomy [41], subtly ask the user whether it should increase its speed.

However, such a classification system will only work properly when it has sufficient input data (i.e. sufficient emotion-data features that are known to distinguish between the states of enjoyment, frustration and boredom) from a sufficient amount of participants. Therefore, the present dataset will be analyzed in more detail, to find more useful features. This will e.g. include the analysis of the video recordings (through facial emotion tracking) and the analysis of in-game data that were gathered during the experiment. The measurements described in Table 1 will also be explored in more detail, by e.g. analyzing standard deviation, variance, kurtosis and skewness of the data (see e.g. [42], [43]). In addition to the *present* physiological state, another possibly useful source of input data lies in the *initial* or *previous* emotional states (e.g. [44], [45]). For example, a high (e.g. 140 BPM) heart rate might not increase as much by an arousing game effect as a normal heart rate. Besides *more* emotion-data features from more participants, also *different* data features are required. In the present experiment, the participants could only report on their gameplay experiences during two rather extreme speed modes (very slow and very fast). Since an emotionally adaptive game should undertake action as soon as the enjoyment borders (Fig. 1-left panel) are crossed, accurate emotion-data on these crossing-border events will be needed.

Acknowledgments

The authors would like to thank Gert-Jan de Vries, Jack van den Eerenbeemd, Paul Lemmens and Floris Crompvoets of Philips Research Laboratories for their contributions to the project.

References

1. Adams, E., Rollings, A.: Game Design and Development; Fundamentals of Game Design. Pearson Education, NJ (2007)
2. Malone, T.W.: What Makes Computer Games Fun? Byte 6, 258–277 (1981)
3. Koster, R.: Theory of Fun for Game Design. Paraglyph Press, Phoenix (2004)
4. Saari, T., Ravaja, N., Laarni, J., Kallinen, K., Turpeinen, M.: Towards Emotionally Adapted Games. In: Proceedings of Presence 2004, Valencia, Spain, pp. 182–189 (2004)
5. Pagulayan, R.J., Keeker, K., Wixon, D., Romero, R., Fuller, T.: User-centered design in games. In: Jacko, J., Sears, A. (eds.) Handbook for Human–Computer Interaction in Interactive Systems, pp. 883–906. Lawrence Erlbaum Associates Inc., Mahwah (2002)
6. Bartle, R.A.: Hearts, Clubs, Diamonds, Shades: Players who suit MUDs, `http://www.mud.co.uk/richard/hcds.htm`
7. Sundström, P., Ståhl, A., Höök, K.: eMoto - A User-Centred Approach to Affective Interaction. In: Tao, J., Tan, T., Picard, R.W. (eds.) ACII 2005. LNCS, vol. 3784. Springer, Heidelberg (2005)
8. Picard, R.W.: Affective Computing. MIT Press, Cambridge (1997)
9. Ekman, P.: An Argument for Basic Emotions. Cognition and Emotion 6, 169–200 (1992)

10. Lang, P.J.: The emotion probe: Studies of Motivation and Attention. American Psychologist 50, 372–385 (1995)
11. Polaine, A.: The flow principle in interactivity. In: Proceedings of 2nd Australasian Conference on Interactive Entertainment, Sydney, Australia, pp. 151–158 (2005)
12. Hunicke, R., LeBlanc, M., Zubek, R.: MDA: A Formal Approach to Game Design and Game Research. In: Proceedings of the Challenges in Game AI Workshop. Nineteenth National Conference on Artificial Intelligence, San Jose, CA (2004)
13. Csikszentmihalyi, M.: Finding flow: The Psychology of Engagement with Everyday Life. Basic Books, NY (1997)
14. Ermi, L., Mäyrä, F.: Proceedings of DiGRA 2005, Vancouver, Canada (2005)
15. Schaefer, F., Haarmann, Boucsein, W.: The Usability of Cardiovascular and Electrodermal Measures for Adaptive Automation. In: Westerink, J.H.D.M., Ouwerkerk, M., Overbeek, T.J.M., Pasveer, W.F., De Ruyter, B. (eds.) Probing Experience: From Assessment of User Emotions and Behavior to Development of Products. Philips Research Book Series, vol. 8, pp. 235–243 (2008)
16. Takahashi, M., Tsuyoshi, A., Kuba, O., Yoshikawa, H.: Experimental Study Toward Mutual Adaptive Interface. In: Proceedings of the 3rd IEEE International Conference on Robot and Human Communication, Nagoya, Japan, pp. 271–276 (1994)
17. Rani, P., Sarkar, N., Liu, C.: Maintaining Optimal Challenge in Computer Games Through Real-Time Physiological Feedback. In: Proceedings of the 1st International Conference on Augmented Cognition, Las Vegas, NV (2005)
18. Journey to Wild Divine, `http://www.wilddivine.com`
19. Bersak, D., McDarby, G., Augenblick, N., McDarby, P., McDonnell, D., McDonald, B., Karkun, R.: Intelligent Biofeedback Using an Immersive Competitive Environment. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201. Springer, Heidelberg (2001)
20. BioTetris, http://www.ncsx.com/www/ncs110998/tetris_64.htm
21. Saari, T.: Mind-Based Media and Communications Technologies. How the Form of Information Influences Felt Meaning. Acta Universitatis Tamperensis 834. Tampere University Press, Tampere (2001)
22. Saari, T., Ravaja, N., Turpeinen, M., Kallinen, K.: Emotional Regulation System for Emotionally Adapted Games. In: Proceedings of FuturePlay 2005, Michigan State University, MI (2005)
23. Öhman, A.: The Psychophysiology of Emotion: An Evolutionary-Cognitive Perspective. In: Ackles, P.K., Jennings, J.R., Coles, M.G.H. (eds.) Advances in Psychophysiology, vol. 2, pp. 79–127. JAI Press, Greenwich (1987)
24. Fasel, B., Luettin, J.: Automatic Facial Analysis: A Survey. IDIAP Research Report (IDIAP RR 99-19). Pattern Recognition 36(1), 259–275 (2003)
25. Van den Hoogen, W., IJsselsteijn, W.A., de Kort, Y.A.W., Poels, K.: Towards Real-Time Behavioral Indicators of Player experiences: Pressure patterns and Postural Responses. In: Proceedings of the 6th International Conference on Methods and Techniques in Behavioral Research, Maastricht, The Netherlands (2008)
26. Ravaja, N.: Contributions of Psychophysiology to Media Research: Review and Recommendations. Media Psychology 6, 193–235 (2004)
27. Yannakakis, G.N., Hallam, J., Lund, H.H.: Entertainment Capture through Heart Rate Activity in Physical Interactive Playgrounds. User Modeling and User-Adapted Interaction, Special Issue on Affective Modeling and Adaptation 18(1-2), 207–243 (2008)
28. Vicente, K.J., Thornton, D.C., Moray, N.: Spectral Analysis of Sinus Arrhythmia: A Measure of Mental Effort. Human Factors 29(2), 171–182 (1987)

29. Veltman, J.A., Gaillard, A.W.K.: Indices of Mental Workload in a Complex Task Environment. Neuropsychobiology 28, 72–75 (1993)
30. Wientjes, C.J.E.: Respiration in Psychology: Methods and Applications. Biological Psychology 34, 179–204
31. Mahlke, S., Minge, M., Thüring, M.: Measuring Multiple Components of Emotions in Interactive Contexts. In: Extended Abstracts of CHI 2006, Montréal, Canada (2006)
32. Partala, T.: Affective Information in Human-Computer Interaction. Academic Dissertation, Department of Computer Sciences, University of Tampere, Finland (2005)
33. Hazlett, R.L., Benedek, J.: Measuring emotional valence to understand the user's experience of software. Int. J. Hum.-Comput. Stud. 65(4), 306–314 (2007)
34. Luczak, H., Göbel, M.: Signal Processing and Analysis in Application. In: Backs, R.W., Boucsein, W. (eds.) Engineering Psychophysiology: Issues and Applications, pp. 79–110. Lawrence Erlbaum, Mahwah (2000)
35. Lacey, J.I., Lacey, B.C.: Verification and extension of the principle of autonomic response-stereotypy. Am. J. Psychol. 71, 50–73 (1958)
36. Mandryke, R.L., Atkins, M.S.: A Fuzzy Physiological Approach for Continuously Modeling Emotion During Interaction with Play Technologies. International Journal of Human-Computer Studies 65, 329–347 (2007)
37. Thayer, J.F., Friedman, B.H.: The Design and Analysis of Experiments in Engineering Psychophysiology. In: Backs, R.W., Boucsein, W. (eds.) Engineering Psychophysiology: Issues and Applications, pp. 59–78. Lawrence Erlbaum, Mahwah (2000)
38. Overmars, M., McQuown, B.: Pacman.GM6 [computer software]
39. Yannakakis, G.N., Hallam, J.: Towards Optimizing Entertainment in Computer Games. Applied Artificial Intelligence 21, 933–971 (2007)
40. Waterink, W.: Facial Muscle Activity as an Index of Energy Mobilization During Processing of Information: An EMG Study. Doctoral Thesis, Tilburg University, Tilburg, The Netherlands (1997)
41. Höök, K., Ståhl, A., Sundström, P., Laaksolahti, J.: Interactional Empowerment. In: Proceedings of CHI 2008, Florence, Italy, pp. 647–656. ACM Press, New York (2008)
42. Picard, R., Vyzas, E., Healey, J.: Toward Machine Emotional Intelligence: Analysis of Affective Physiological State. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(10), 1175–1191 (2001)
43. Van den Broek, E.L., Schut, M.H., Westerink, J.H.D.M., Van Herk, J., Tuinenbreijer, K.: Computing Emotions Awareness through Facial Electromyography. In: Huang, T.S., Sebe, N., Lew, M., Pavlović, V., Kölsch, M., Galata, A., Kisačanin, B. (eds.) ECCV 2006 Workshop on HCI. LNCS, vol. 3979(12), pp. 52–63. Springer, Heidelberg (2006)
44. Wilder, J.: Stimulus and Response: The Law of the Initial Value. J Wright, Bristol, England (1967)
45. Berntson, G.G., Cacioppo, J.T., Quigley, K.S.: Autonomic Determinism: The Modes of Autonomic Control, the Doctrine of Autonomic Space, and the Laws of Autonomic Constraint. Psychological Review 98(4), 459–487 (1991)

Robust Interactive Storytelling Framework for Automatic TV Content/Story Production

Hussein Karam Hussein Abd El-Sattar

Ain Shams University, Faculty of Science, Mathematics and Computer Science Department, Abbassia 11566, Cairo, Egypt
huss20@hotmail.com

Abstract. With the rapid advancement of hardware and software, entertainment computing industry has been popularized during the last decade. TV content/story production is a major bottleneck in the entertainment industry. Previous work has identified key problems, such as narrative control, the duality between character and plot, the tension between interactivity and storytelling, and the tools that artists can use to create interactive story worlds. Due to the tradeoff between narrative progression and user interaction, striking a balance between the two is a big challenge. As a solution to such a challenge, Bcreative system has been developed. It is a unified prototype of both plot-based and character-based interactive storytelling. The idea is derived from the study of both Petri nets (PN) and cinematography in interactive storytelling. The system architecture is composed of two engines: a story engine and a character engine through a smart user-friendly graphical user interface (GUI). The story engine models the story plot based on the Petri nets (PN) reachability analysis, while the character engine enhances the modeling of character behavior based on filmmaking theory. This paper discusses the development process of our proposed system, illustrating the design and implementation in detail. Comparisons with existing techniques are included, and the advantages of our proposed system over existing ones are examined. Evaluation and practical performance results are also demonstrated.

1 Introduction

Interactive storytelling (IS) is a promising growth area in computer-based technologies, in the entertainment computer industry, and other applied areas. TV content/story production is a major bottleneck in the entertainment industry. The internet has certainly created a major stir, but the fact remains that TV content/story generation is still the primary means by which most people around the world get their information. It is generally assumed that individuals cannot easily produce TV content/story with conventional techniques because content production greatly consumes time, effort, budgets and it is not only limited to producers, directors, actors/actresses, etc., but also requires some level of explicit narrative representation and narrative control. The plot of the story and the character(s) are the two most important elements of interactive storytelling. In the field of interactive storytelling technology, we try to bridge a connection between player desires and story content to provide a deep connection between what the player does in the story world and where

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 134–148, 2008.

the story leads. Due to the tradeoff between story narrative and user interactions, striking a balance between them as well as the duality between character and plot throws up an aspiring goal and challenging task. As a solution to such a challenge, Bcreative system has been developed. It is a unified prototype of both plot-based and character-based interactive storytelling. The system structure is composed of two engines: a story engine and a character engine through a smart user-friendly graphical user interface. The story engine models the story plot based on the Petri nets (PN) reachability analysis [5,14,18], while the character engine enhances the modeling of character behavior based on filmmaking theory [4,9,6]. The approach used a Petri nets as a state-based model for story plot generation, where each scene in a plot is regarded as a goal to be reached. A complex scene is represented as a composite goal, and a simple scene is represented as an atomic goal.

This paper is organized as follows. Section 2 introduces some terminology and basic notions that will be used. Section 3 outlines the previous relevant work on interactive storytelling technology and the problems that were encountered. The development process of our proposed system including system structure, methodology, design, etc. is discussed in section 4. System configuration is given in section 5. Evaluation and practical performance results with comparisons are discussed in section 6. Section 7 concludes this paper and outlines some directions for future work.

2 Background and Basic Notions

For clarity and accuracy, this section summarizes the background and the basic terminology and notions that will be used throughout this paper.

2.1 Interactive Storytelling Background

Interactive storytelling is a developing kind of computer entertainment in which the player plays the role of the protagonist in a dramatically rich environment. In traditional forms of storytelling, a storyteller would present the scenario of a story to the audiences in a predefined way (also known as a plot), which limited the variation in character interactions and context. The story refers to the succession of actions that happen in the world represented by the narrative. A narrative means a certain type of artistic and social expression, where a kind of imitation of real events is involved. In dramatic writing, stories are thought of as consisting of events that turn (change) values. A value is a property of an individual or relationship, such as trust, hope, etc. Semantically, a scene (story event) is an atom of a story plot, which involves the interactions among a number of virtual actors with dedicated roles, under a certain context. An actor is the virtual entity which performs the tasks assigned from the drama manager (director). Generally, scenes are composed of beats, where the term beat refers to the smallest unit of action that has its own complete shape. A beat is a dramatic action that occurs in a scene to achieve a narrative goal. A beat consists of (1) preconditions, a list of predicates that need to be true for the beat to be selected; (2) post conditions; a list of predicates that will be true as a consequence of firing the beat; (3) success and failure conditions, and (4) a joint plan to be executed by the actors. Beats are the fundamental unit of actor guidance and it defines the granularity of plot/actor interaction. It serves several functions and plays an important role within the structure of our proposed interactive storytelling framework.

2.2 Cinematography and Interactive Drama Technology

Drama is a special kind of narrative where the actions are directly represented to the spectator. There have been three main approaches in creating interactive drama technology: “Branching Narrative (Lots of story, little interactivity)”, “Narrative Game (Lots of interactivity, little story)”, and “Storytronics (Lots of both story and interactivity)”. In Storytronics technology, interactive storytelling is achieved through artistic works called story worlds. A story world is a universe of dramatic possibilities. A story world is played by a player, who controls the protagonist actor in that story, with the other actors controlled by the computer. It is comprised of two groups of information: (i) dramatic components, and (ii) dramatic principles. Dramatic components are the same things that make up any story - the story’s actors, locations, props, and occurrences. Dramatic principles are the unique way in which a story builder expresses his or her authorial voice. A story world is built up of many kinds of elements. These elements are divided into three broad categories: Deikto words, behavior controllers, and scripts. Behavior controllers are story world elements which are invisible to the player, but which affect the way story world actors behave. Deikto is a form of simplified English used by the storyteller. Deikto words are story world elements which the player can come in direct contact with. A Deikto sentence describes a dramatic action taken by an actor in the story world. If past, it is called an event, if planned for the future it is called a plan. Storyworlds contains all the concrete pieces that are used to build up the story such as storyworld’s actors, places, stages, props as well as past and present occurrences and future plans. On the other hand, scripts are applied by actors at certain times, causing changes in the storyworld or directing actors on how to behave. Scripts are created using a special scripting language. This language is built out of operators, the same way that human language is built out of words. Bcreative uses a modified TVML as a cinematic and scripting language [6].

2.3 Petri Nets (PN)

A Petri nets (PN) [14,18] is an abstract, formal model of information flow. It is a graphical and mathematical modeling tool applicable to many systems. As a graphical tool, PN can be used as a visual-communication aid similar to flow charts, block diagrams, and networks. As a mathematical tool, it is possible to set up state equations, algebraic equations, and other mathematical models governing the behavior of systems. A Petri nets $PN = (N, M_0)$ consists of a structure N and an initial marking M_0, where: N=(P,T,F,W) is a Petri nets structure,

$P = \{p_1, p_2, \cdots, p_m\}$ is a finite set of m places,

$T = \{t_1, t_2, \cdots, t_n\}$ is a finite set of n transitions,

$F \subseteq (P \times T) \cup (T \times P)$ is a set of arcs, where $P \cap T = \phi, P \cup T \neq \phi.$

$W : F \to \{1,2,3,\cdots\}$ is a mapping which associates to each arc of the net its weight,

$M_0 : P \to \{1,2,3,\cdots\}$ is the initial marking representing the initial state of PN.

A special (PN) in which place capacities and arc weights are equal to one is called a condition/event net (C/E net). A transition without any input place is called a source transition and one without any output place is called a sink. Labels for unitary weight are usually omitted. In modeling using PN, we regard the places as conditions and the transitions as events. On the other hand, the state of PN is described by means of the concepts of marking. A marking is a function that assigns to each place a nonnegative integer called a token. A token is a primitive concept of PN like places and transitions. From a graphic point of view, places are usually represented by circles, transitions by rectangles and marks by black dots into places. The dynamics of the net is described by moving tokens among places according to the following transition firing rules:

1. A transition t is said to be enabled if each input place p of t is marked with at least W(p,t) tokens, where W(p,t) is the weight of the arc from p to t.
2. An enabled transition may or may not fire depending on whether or not the event actually takes place.
3. A firing of an enabled transition t removes W(p,t) tokens from each input place p of t and adds W(p,t) tokens to each output place of t, where W(p,t) is the weight of the arc from p to t.
4. The marking of other places which are neither inputs nor outputs of t remains unchanged.

3 Problem Statement and Prior Research

3.1 Problem Statement

Story plot and character(s) are the two most important elements of a story. Based on these two elements there are currently two research directions: plot-based and character-based interactive storytelling. As interactive storytelling is a mixture of conventional storytelling and user interactions, and due to the tradeoff between story narrative and user interactions, striking a balance between the two is a big challenge. Consequently, there is a need to develop an integrated framework to achieve the balance between conveying a story's moral and enhancing the modeling of character behaviors. As a solution to such a challenge, Bcreative system was developed. It is a unified prototype of both plot-based and character-based interactive storytelling. It gives the opportunity for the user to get all the ideas out of his mind and be creative, hence the name Bcreative. This occurs in a short time, in a simple way with great efficiency, and with no budget by manipulating the embedded smart user-friendly graphical user interface. The idea of Bcreative is derived from the study of Petri nets, cinematography in interactive storytelling, and filmmaking theory.

3.2 Prior Research

Recent developments in interactive storytelling have resulted in many different approaches which differ on various dimensions and sometimes overlap, such as: immersive storytelling [12]; emergent storytelling [1]; interactive story authoring/generation [10]; plot-based [7,13,17] and character-based interactive storytelling [2,3]. Pérez y Pérez, R. & Sharples, M. [16] present an updated review of different interactive storytelling systems with comparisons. More recently, a hybrid system of plot-based and character-based interactive storytelling is discussed in [19]. One of the characteristics of the previous approaches is in controlling a balance between interaction and narration. They showed a lot of interesting results. However, they didn't consider a user's level of understanding, even though people have different abilities to understand a story according to their knowledge, experience, age, gender, etc. Moreover, previous work has identified relevant dimensions and key problems for the implementation of interactive storytelling, among which are: the status of the user, the level of explicit narrative representation and narrative control, the modes of user intervention, the duality between character and plot, the tension between interactivity and storytelling, etc. This paper attempts to address these problems, in particular, the relations between character and plot by modeling the story plot as well as the character behavior in an integrated framework through a smart-user friendly graphical user interface (GUI) platform called Bcreative.

4 Proposed System Development Process

Developing an interactive storytelling system to achieve the autonomy of story generation and story performing is an exhaustive and complicated process requiring careful design and layout of the story content and its visual presentation. This section discusses the development process of our proposed system from the earlier design phase called the process model up to the final product system.

4.1 Methodology and System Workflow

The methodology in our system is based on a basic ontology which holds the various concepts that are relevant to story plot generation as well as character behavior. The system architecture is a composition of two engines: story engine and character engine through a smart user-friendly graphical user interface. The story engine models the story plot based on the PN reachability analysis [5], while the character engine enhances the modeling of character behavior based on cinematography in interactive storytelling, and the story world elements (e.g. actors, camera, lighting, etc.). The system workflow is shown in Fig. 1, where the content inside the oval shape represents an output from the previous task or state.

Based on the circumstances that a story scene occurs (story context), Fig. 1 shows that the TV content/story is produced from a set of various meaningful storylines or plots, each of which is composed of scenes, which are modeled based on PN reachability, and will be explained in detail in subsection 4.3. In our approach, we used Petri nets (PN) as a state-based model for story plot, where each scene in a plot is regarded as a goal to be reached. A complex scene is represented as a composite goal, and a simple scene is represented as an atomic goal. The director (Bcreative user) acts as the storyteller, who dynamically selects and determines the storyline from the whole plot based on the user interaction or context variables. A scene of the storyline is composed of beats. A beat is a dramatic action or the smallest story unit that occurs in a scene to achieve a narrative goal. A beat in our system consists of an action/reaction pair between actors and will be generated in real-time using a character

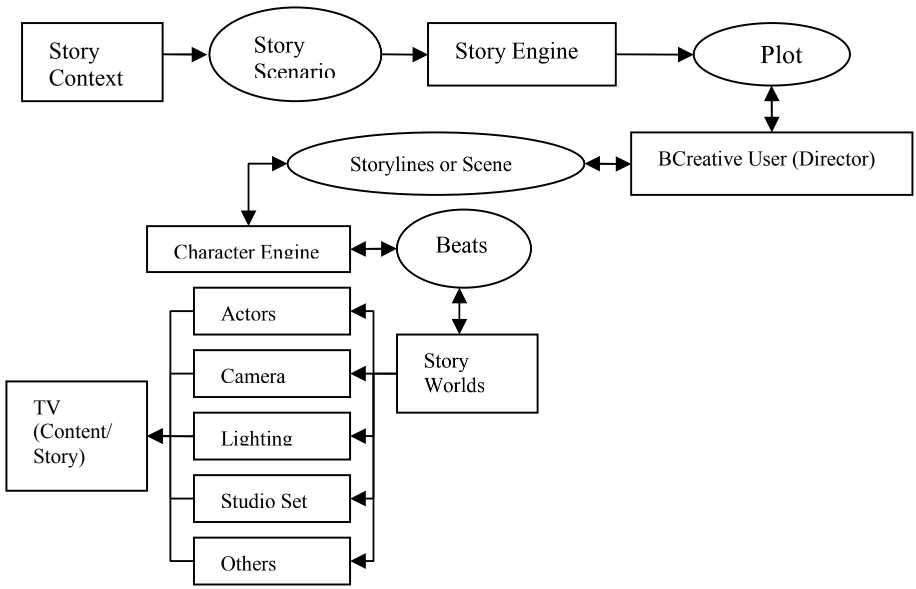


Fig. 1. System Workflow

engine and executed through artistic works called story worlds via its dramatic components (e.g. actors, camera, etc.). For a complex scene in the presentation path, the goal can be decomposed into more specific sub-goals. Depending on the user interactions and context, different consequent goals may be reached after a certain goal, i.e. different scenes are achieved in different situations.

4.2 Architectural Design and Graphical User Interface (GUI)

Design involves decomposing the system into modules. The result is a design specification document which contains a description of the software architecture: what each

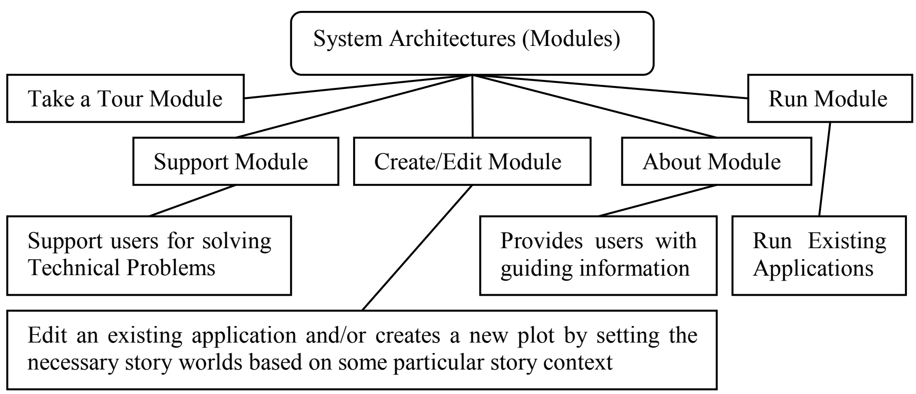


Fig. 2. Architectural Design

module is intended to do and the relationships among modules. Fig. 2 shows some of the basic architectural design of our proposed storytelling framework. Fig. 3 shows these different modules via the main graphical user interface (GUI) screen window once we run the Bcreative system.

Fig. 3. Bcreative Main (GUI) Screenshot

4.3 Story Engine

Reachability is the fundamental basis for studying the dynamic properties of any system [5]. The firing of an enabled transition will change the token distribution in a Petri nets according to the transition firing rules described in subsection 2.3. The idea behind using Petri nets (PN) to model the story plot as well as character behavior is based on three factors: (i) the story is generated from a set of smaller sub-stories or plots, each of which is composed of scenes, and scenes are composed of beats, (ii) the degree of freedom, which the PN possesses (e.g. marking of the net, the sequence of firing, etc.) as well as PN properties such as, sequential execution, conflict, concurrency, confusion, inhibitor arcs to model priorities, etc., and (iii) the concept of PN reachability which can be simply stated as "Can we reach one particular state from another?". Based on these factors, our idea is derived for using PN as a state-based model for story representation/plot. Each scene in a plot is represented as a graph of desired states to be reached. The final goal states for a scene mark the end of the scene, just as the initial state describes how the scene should begin. A state is marked with a transition from one state to another, which provides a relative ordering of how states should flow temporally. Transitions are "source-action-target" triplets where the execution of an action triggers a change of state. For the state of a Petri nets only the set of places P is relevant, because the network structure of a Petri nets does not change and only the distribution of tokens over places changes. A state, also referred to as marking, corresponds to a mapping from places to natural numbers. Marking is based on the number of beats the scene constitutes, and plots are constructed based on the firing the beat, where a scene of the story is regarded as a goal. Any state s can be presented as $s \in P \rightarrow \{0,1,2,\ldots\}$, i.e., a state can be considered as a multi-set, function, or vector. In the context of state spaces, we use places as attributes. In any state the value of each place attribute is known: $s(p)$ is the value of attribute $p \in P$ in state s. Since, firing a transition will change the state marking M to a new marking M'. The state space of a Petri nets with n places is the set of all markings, that is N^n. Now the question that arises is how to decide which scene to attempt to make happen next? For this, we should examine the list of unused scene and choose the one that has a satisfied precondition and whose value change best matches the shape of the global plot arc. We formulated the state change caused by a transition firing using a partial function δ which we call the next-state function, where,

$$\delta : N^n \times T \rightarrow N^n \tag{1}$$

The function δ when applied to a marking M and a transition t_j yields the new marking M' which results from firing the enabled transition t_j in the marking M, that is,

$$\delta(M, t_j) = \begin{cases} undefined & if \quad t_j \\ & is \quad not \quad enabled \\ M' & if \quad t_j \\ & is \qquad enabled \end{cases} \tag{2}$$

M' is the marking which results from removing tokens from the inputs of t_j and adding tokens to the output of t_j. The function δ given by equation (2) incorporates a notion of distributed state, and a rule for state change of a Petri net via a sequence of markings $(M_0, M_1, M_2, \ldots)$ and a sequence of transitions $(t_{j(0)}, t_{j(1)}, t_{j(2)}, \ldots)$ which were fired. The relationship between these two sequences as well as the properties of Petri nets [18] such as, sequential execution, conflict, concurrency, confusion, duality, conflict, zero-testing through the introduction of inhibitor arcs to model priorities etc., form the main idea for story engine to model the story plot. An inhibitor arc connects a place to a transition and is represented by a dashed line terminating with a small circle instead of an arrowhead at the transition. The introduction of inhibitor arc adds the ability to test “zero” (i.e. absence of tokens in a place). The notion of conflict, on the other hand, is often described as the main mechanism of drama, where it occurs when a possible action for a character is not compatible with his/her value, e.g. a solitary boy must approach a girl, an aristocratic lady can’t reveal her love for a poor servant, etc. Fig. 4(a-e) demonstrates some of the (PN) properties such as sequential execution, conflict, concurrency, confusion, etc. which are represented using a set of simple constructs as follows:

A. Sequential execution: (Fig. 4(a)), transition t_2 can fire only after the firing of t_1. This imposes the precedence of constraints “t_2 after t_1”.
B. Conflict: Transitions t_1, t_2 and t_3 are in conflict as shown in Figure 4(b). All are enabled but the firing of any leads to the disabling of the other transitions.

C. Concurrency: (Fig. 4(c)), the transitions t_1, t_2, and t_3 are concurrent. A necessary condition for transitions to be concurrent is the existence of a forking transition that deposits a token in two or more output places.

D. Priorities: (Fig. 4(d)), the classical Petri nets have no mechanism to represent priorities. Inhibitor nets defined special arcs called inhibitor arcs to model priorities.

E. Confusion: A situation where concurrency and conflicts co-exist (Fig. 4(e)). Both t_1 and t_3 are concurrent, while t_1 and t_2, are in conflict. Also, t_2 and t_3 are in conflict.

Equation (2) can be written in a compact recursive manner as follows:

$$\delta\left(M_k, t_{j(k)}\right) = M_{k+1} \text{ for } k=0,1,2, \tag{3}$$

Equation (3) forms an iterative process of marking which exhibits a certain analogy of the recursive process for story plot modeling. Fig. 5 shows a sample story scenario example for equation (3), which contains scenes from P_1 (start state) to P_4 (final or goal state) with initial marking $M_0 = (1,0,1,0,2)$.

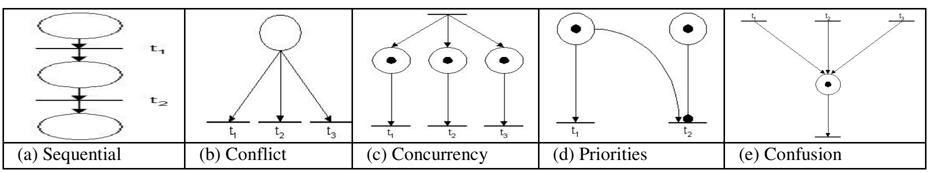


Fig. 4. Some PN Properties illustrations

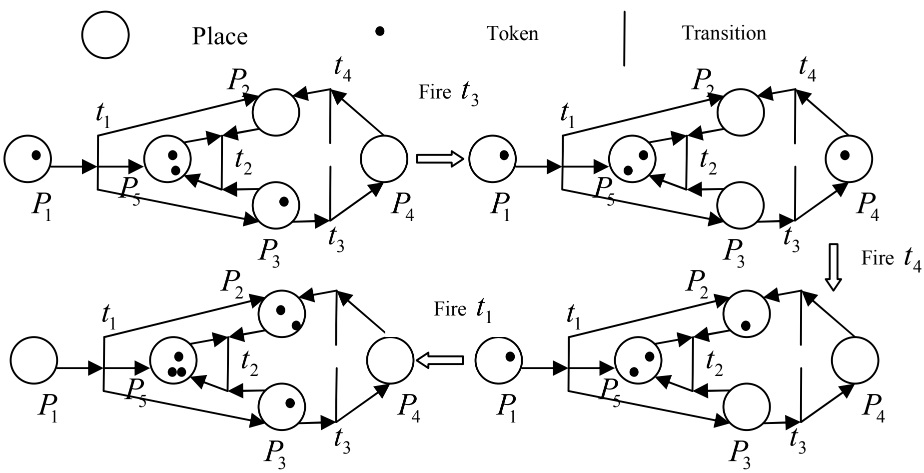


Fig. 5. Scene Scenarios Example with Marking Reachability

In Fig. 5, tokens are regarded as beats where marking is based on. For instance, with the initial marking $M_0 = (1,0,1,0,2)$, two transitions t_1 and t_3 are enabled, choosing t_3 producing the marking $\delta(M_0, t_3) = (1,0,0,1,2) = M_1$. In this marking, transitions t_1 and t_4 are enabled, firing t_4 changes the marking to $\delta(M_1, t_4) = (1,1,0,0,2) = M_2$. In M_2, t_1 is enabled and firing it leads to $\delta(M_2, t_1) = (0,2,1,0,3) = M_3$.

4.4 Character Engine

Based on the user interactions or context, the director (Bcreative user) dynamically selects and determines the scenes from the whole plot obtained previously from story engine, and then assigns the scene contents folded by its dramatic components (e.g. camera, lighting, etc.) to different actors involved in the scene via the character engine and based on filmmaking theory. Story plot and character interactions are two important elements of interactive storytelling. However, we believe that interactive stories also need to include investigation dealing with the visual presentation of the story and the way it is shown and told. Theatre directors, filmmakers, and animators have emphasized the importance of visual design. Film and theatre, as well as animation artists, spend many hours, days, or even months, creating a visual design for a production. A visual design is considered the heart of a performance; it deepens and enriches the dramatic experience through artistic works called story worlds via its dramatic components which have direct impact on communicating the narrative, evoking emotions and moods, and engaging viewers. These components have many psychological and aesthetic effects that influence viewers' perception of a scene. We considered these important factors in developing our system and modeled the process of the character engine according to program-production processes used in the real world. Character engine deals with the visual presentation of the story and the way it is shown and told, in the context of applying a formal cinematic language to the visual components, in order to give a better shape to the story, and to the whole dramatic experience [4,9]. Cinematic language, translated from a movie to a virtual environment, should be a valid option, with different proved formalisms that can be translated to a storytelling system. Our system embedded a modified TVML [6] as a cinematic language for an action/reaction pair between actors, executed through story worlds via its dramatic components in order to provide better dramatic experiences and supporting the narrative relevance of the stories being narrated. The character engine structure follows standard formats for content production which is classified into two categories: "information data" and "presentation data". Information data refers to the information that is to be conveyed to viewers via the program. Presentation data describes the way in which that information should be shown to viewers as a program. Presentation data is the core of any content production which needs some resources or production modules obtained from the information data. Character engine contains two modes: scene mode and editing mode. Both of them are supported by a user-friendly graphical user interface (GUI) screens. Screenshots for some of these screens are shown in Fig. 6. Each screen supports level of details (LOD) rendering techniques. The user starts in scene mode to execute his/her scene constructions (tasks) required for creating the desired shots based on the information data and/or the story plot given. The user then moves to editing mode to edit, run, manipulate, play movies or preview the shots which have been previously constructed. Character engine architecture is shown in Fig. 7.

Fig. 6. Screen shoots of character engine GUI supporting screens

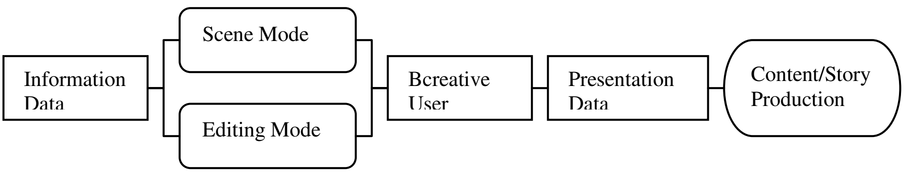


Fig. 7. Character Engine Structure

5 System Configuration

Bcreative environment is a general purpose TV content/story production application system. It is compatible with different Windows operating systems. The hardware environment requires computers with multimedia support. The software environment is based on Microsoft Visual studio tools, Microsoft DirectX 9 drivers, TTS (Text to Speech) engine, and the multimedia tool called TVML [6]. TVML is a scripting language used for story world and embedded in our system. Unfortunately, TVML involves a limited number of characters; therefore, the system may allocate one character to multiple actors, which confuses the viewers. In our implementation, we solved this problem by increasing the number of characters as well as probes usable in TVML. The Bcreative user doesn't need to understand the TVML (more than 90 commands) because our system supports the user with a smart user-friendly screens.

6 Evaluation and Performance Results

6.1 User Test

The functionality and the performance of the proposed system have been evaluated according to several criteria such as: (i) the creativity process including psychological and historical creativity, story-predictability including structure and content elements, (ii) story-development processes and knowledge data-bases stored by the system, including rhetoric-knowledge, story-world knowledge and common-sense knowledge, (iii) interestingness which is a combination of content, originality, suspense and coherence (the user's ability to comprehend the relationships between the events in the story, both within the story world and storytelling), and (iv) use of the natural language structure and meaning. These criteria are the elements of literary criticism that cannot be captured by objective measures. The system was shown to a few visual artists and designers who are practically test the system and reported the following results obtained for the story (in a scale from 1 to 5, where 5 indicates best): narrative flow: 3.9; clever plot: 4.1; coherency: 4.5; use of language: 4.4; content: 3.9; suspense: 4.2; creativity: 4.7. They showed great interest in the system, especially in its use as a rapid prototyping tool for writers. They reported that the system has shown great utility as an authoring tool for content production and can be objectively used as a teaching aid for many purposes like education, cinematography, communication, etc. Comparisons with existing systems showed that the Bcreative system process is similar to the Minstrel model [16] in knowledge structures where they are entered by a graphical interface that allows dropping frames from a menu and interconnecting them with relationships. Similarly to the Mexica model [15]; its story-predictability can be classified as low. In addition, it ensures novelty most like both Minstrel and Mexica models. Table 1 compares the Bcreative system with some other existing systems like Façade [11], I-storytelling [20], and hybrid system [19].

Table 1. System Comparisons with other system models

Name / Factor	Facade	I-Storytelling	Hybrid	Bcreative
Description	Credible characters controlled by a dramatic manager.	Virtual actors with dramatic goals	Multi-agent storytelling system	An authoring environment for virtual storytelling
Methodology	ABL programming, beat, dramatic arch.	Hierarchical Task Network (HTN) planning	Fuzzy Cognitive Goal Net (FCGN)	Petri nets, filmmaking theory and cinematography in interactive storytelling
Technology	Java & JESS reasoning.	Unreal and C++ planning	Active Worlds 3-D virtual environment powered by Renderware	Microsoft visual studio tools, TVML, and other multimedia tools
Application	An interactive drama.	Interactive television	Interactive storytelling	Interactive TV content/story production
Scalability	Intermediate	Intermediate	Intermediate	High
Replayability	High	High	Intermediate	High
Creativity	High	Intermediate	Intermediate	High
Natural lang. production	High	Intermediate	Low	High
Graphical user Interface (GUI)	Intermediate	Intermediate	Low	High
Interestingness	High	High	High	High

6.2 Case Study

One of several real case studies which have been done for testing our system is based on a story scenario of horse riding (Endurance & Jumping). Horse riding is an interesting and famous sport worldwide [8]. We tested our system for producing TV content/story of horse riding as a teaching aid for beginners. Fig. 8 shows the screen shots for such content production. The corresponding Petri nets used to model the story scenario is shown in Fig. 9.

Fig. 8. Horse riding content/story production example created by our System

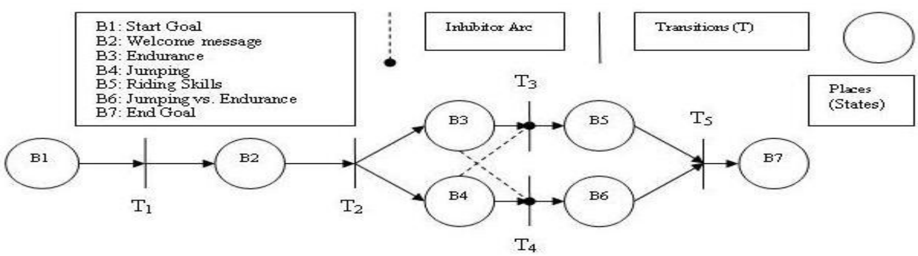


Fig. 9. Corresponding PN for scene scenario of figure 7 with beats B

7 Conclusion and Future Work

TV content/story production is a major bottleneck in the entertainment industry. In this paper, we have proposed an authoring environment for virtual storytelling which attempts to reconcile the tension between interactivity and storytelling as well as narrative progression and user interaction. The developed system is composed of two engines: a story engine and a character engine through a smart user-friendly graphical user interface. The story engine models the story plot based on the PN reachability analysis, while the character engine enhances the modeling of character behavior based on filmmaking theory. The proposed system can be objectively used as a teaching aid for many purposes like education, cinematography, communication, etc. The merits and the original contribution of this paper are:

1) Presenting a new idea for interactive storytelling technology using a new kind of analysis method for the marking of the PN and its reachability analysis. We show that PN as a graphical and modeling tool can be used as a visual-communication aid in films to model the computer graphics scenes (story plot) and characters.
2) Reporting that most computer-based application techniques developed outside the realm of Petri nets can be applied to Petri nets as well.
3) Developing an integrated framework of both plot-based and character-based interactive storytelling to achieve the balance between conveying a story and enhancing the modeling of character behaviors as well as reconcile the character-based approach with the problem of narrative control.
4) Enhancing scalability issues, which are satisfied by distinguishing the story beats and their visual presentation, where this abstraction has enhanced scalability and promoted reuse by enabling the reuse of the same visual presentation with different story beats.
5) Supporting replayability (entertainment value), the creativity process and others.

Creativity is a fascinating area where there are too many things waiting to be discovered. Hopefully, this paper will encourage further work on computer models of creativity. Despite the great efficiency of the obtained results, currently we are working on extending our system to be applicable in multi-language with speed animation enhancement.

References

[1] Aylett, R.: Narrative in Virtual Environments: towards emergent narrative. In: AAAI 1999 Fall Symposium on Narrative Intelligence (1999)

[2] Cavazza, M., Charles, F., Mead, S.J.: Interacting with Virtual Characters in Interactive Storytelling. In: ACM Joint Conference on Autonomous Agents and Multi-Agent Systems, Bologna, Italy, pp. 318–325 (2002)

[3] Cavazza, M., Charles, F., Mead, S.J.: Character-based Interactive Storytelling. IEEE Intelligent Systems, special issue on AI in Interactive Entertainment, pp. 17–24 (2002)

[4] Courty, N., Lamarche, F., Donikian, S., Marchand, E.: A cinematography system for virtual storytelling. In: Balet, O., Subsol, G., Torguet, P. (eds.) ICVS 2003. LNCS, vol. 2897, pp. 30–34. Springer, Heidelberg (2003)

[5] Huang, J., Tadao, M.: Classification of Petri net transitions and their application to firing sequence and reachability problems. In: IEEE Conference on Systems, Man, and Cybernetics, vol. 1, pp. 263–268 (1997)

[6] Hayashi, M., Douke, M., Ariyasu, K., Hamaguchi, N.: TV program Making language - Benefits of object-based contents-creation method. In: IBC 2001 Conf., pp. 23–31 (2001)

[7] Hussein, H.K.: A new framework for Plot-based interactive storytelling generation. In: 5th IEEE International Conference Computer Graphics, Imaging and Visualization, Malaysia, Pennang, August 25-28 (accepted) (to appear, 2008)

[8] Kijima, R., Kouno, M., Hashimoto, K., Jiang, Y., Aoki, T., Ojika, T.: Karakuri Horse Riding Therapy. In: Proc. of the Int. Conf. on Rehabilitation Robotics, pp. 278–281 (2003)

[9] El-Nasr, M.S.: Interactive narrative architecture based on filmmaking theory. International Journal on Intelligent Games and Simulation 3(1) (2004)

[10] Magerko, B., Laird, J.: Building interactive drama architecture. In: First Int. Conf. on Technologies for Interactive Digital Storytelling and Entertainment, pp. 226–237 (2003)

[11] Mateas, M., Stern, A.: Façade: An experiment in building a fully-realized Interactive Drama. In: Game Developers Conference, Game Design track (2003)

[12] Nakatsu, R., Tosa, N.: Interactive Movies. Handbook of Internet and Multimedia. IEEE, Los Alamitos (2000)

[13] Pablo, G., Díaz-Agudo, B., Federico, P., Raquel, H.: Story plot generation based on CBR. Knowledge-Based Systems Journal 18, 235–242 (2005)

[14] Peterson, J.L.: Petri nets. ACM Computing Surveys 9(3), 223–252 (1977)

[15] Perez y Perez, R., Sharples, M.: MEXICA: A computer model of a cognitive account of creative writing. Journal of Experimental and Theoretical AI (2001)

[16] Perez y Perez, R., Sharples, M.: Three computer-based models of storytelling: BRUTUS, MINSTREL and MEXICA. Knowledge Based Systems Journal 17(1) (2004)

[17] Sgouros, N., Papakonstantinou, G., Tsanakas, P.: A Framework for Plot Control in interactive story systems. In: Proc. Am. Assoc. Artificial Intelligence (AAAI 1996), pp. 162–167. AAAI Press, Menlo Park (1996)

[18] Tado, M.: Petri nets: properties, analysis and applications. Proc. IEEE 77 (1990)

[19] Cai, Y., Miao, C., Tan, A.-H., Shen, Z.: A Hybrid of Plot-Based and Character-Based Interactive Storytelling. In: Hui, K.-c., Pan, Z., Chung, R.C.-k., Wang, C.C.L., Jin, X., Göbel, S., Li, E.C.-L. (eds.) EDUTAINMENT 2007. LNCS, vol. 4469, pp. 260–273. Springer, Heidelberg (2007)

[20] I-storytelling, `http://www-scm.tees.ac.uk/users/f.charles/`

Designing Toys That Come Alive: Curious Robots for Creative Play

Kathryn Merrick

School of Information Technologies and Electrical Engineering
University of New South Wales, Australian Defence Force Academy
Northcott Drive, Canberra ACT, 2612
k.merrick@adfa.edu.au

Abstract. Creative thinking requires imagination, creativity, play, sharing and reflection. This paper presents an architecture for a curious, reconfigurable robot that encourages creative design thinking by permitting designed structures to learn behaviours. These behaviours encourage designers to play with different structures, reflect on the relationship between structure and behaviour and imagine new structures. A demonstration of the architecture is described using the Lego Mindstorms platform. The demonstration shows how a curious robot can adapt new behaviours in response to changes in its structure, and how this can encourage the creative thinking spiral and creative design.

Keywords: Curiosity, motivated reinforcement learning, robots, creative design, creative play.

1 Creativity and Design

Design can be described as a process of purposeful, constrained decision making requiring exploration and learning [2]. The role of creativity in design has many interpretations, including the distinction between the design of creative artefacts and the evaluation of the processes involved in design as creative [1, 3, 10]. Renick [7] writes that there has been a transition from the Industrial Society to the Information Society and, most recently, the Creative Society. He identifies the need to teach people to use creative thinking processes involving imagination, creativity, play, sharing and reflection.

This paper presents an architecture for curious, reconfigurable robots as a toy that encourages creative play. The architecture permits designed structures to learn behaviours to encourage play, reflection and imagination for ongoing, creative design.

The remainder of this section discusses reconfigurable robots as toys and introduces the idea of curious robots to encourage creative play. Section 2 presents an architecture for curious, reconfigurable robots using the Lego Mindstorms platform. Section 3 describes an initial demonstration of a curious robot learning behaviour in response to changes in its structure. The paper concludes by reflecting on how curious robots can encourage creative design by influencing the creative thinking spiral [7].

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 149–154, 2008.

1.1 Reconfigurable Robots and Creative Play

Reconfigurable robots comprise sets of modules that can be re-arranged to achieve different structures, behaviours and functions. Reconfigurable robots have been developed for their functional, economic and creative advantages. Functionally, reconfigurable robots promise engineering versatility, flexibility, robustness and the ability to self-repair through redundancy [13]. Likewise, economic advantage can be gained by designing complex machines as reconfigurable sets of modules.

From a creative viewpoint, technologies for reconfigurable robots include toys such as Lego (http://www.lego.com) and Meccano (http://www.meccano.com). These products encourage creative play, creative thinking and creative design by providing sets of basic components and connector modules that can be built into different structures. Recent versions of these toys can be fitted with motorised and even electronic components to control the behaviour of the final structure. Meccano creations can be fitted with motors or remote controls and Lego Mindstorms products can be attributed behaviour using a programmable brick.

The incorporation of motors, remote controls and programmable bricks opens the way traditional building packages such as Lego and Meccano to become platforms for playful learning of concepts in electronics and computer programming. In contrast, this paper proposes an architecture that draws on the programmable capacity of reconfigurable robots to encourage creative design thinking through creative play. This architecture uses computational models of curiosity and machine learning to create curious robots that are able to 'come alive' with new behaviours in response to changes in their structure. The aim is to encourage experimentation with structure by providing the creative designer with real-time information about the relationship between the structure and its possible behaviours.

1.2 Curious Robots

While existing research has focused on developing hardware modules for reconfigurable robots, and the software required for those modules to communicate, the problem of attributing behaviour to reconfigurable robots remains a challenge. Traditional artificial intelligence techniques tend to assume a fixed set of inputs and outputs (sensors and actuators) and a fixed set of goals based on these inputs and outputs [8]. This is in contrast to the needs of reconfigurable robots, which may have changing sensors and actuators and thus changing goals.

Recent work has focused on the design of developmental learning algorithms that can generate goals online in response to the changing experiences of a robot [6, 9]. These approaches uses computational models of intelligent, adaptive, curiosity to develop new goals, but do not provide a way for robots to adapt to changes in their sensors or actuators. A later approach that includes adaptable representations for input and output was proposed by Merrick and Maher [5]. This paper adapts the Merrick and Maher [5] model for motivated reinforcement learning agents to design an architecture for curious, reconfigurable robots that can develop new behaviours as a response to their changing structure.

2 Architecture for a Curious, Reconfigurable Robot

The architecture presented in this section comprises two layers, a device layer and an agent layer, as shown in Figure 1. This paper uses the Lego Mindstorms NXT robot platform running the Lejos (http://lejos.sourceforge.net) Java firmware. However the concept of a reconfigurable robot using curiosity and motivated reinforcement learning is general enough to be applied to other reconfigurable robot hardware.

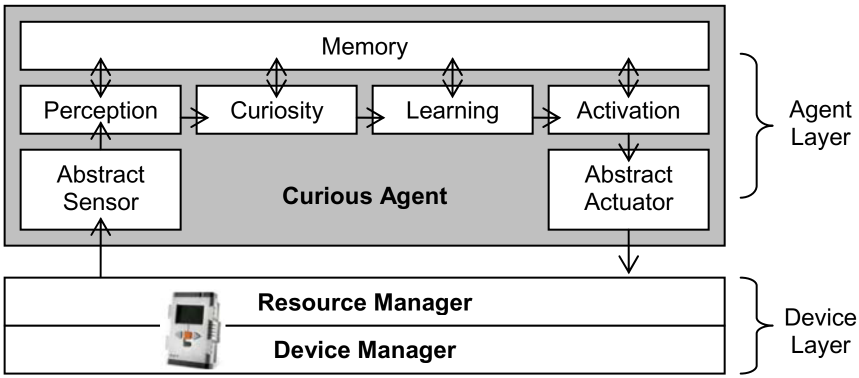


Fig. 1. Architecture for a curious, reconfigurable robot using the Lego Mindstorms platform

2.1 Device Layer

Lego Mindstorms robots can be designed with a range of different sensors and actuators including sensors for colour, light, sound or distance to objects and actuators for servo motors. NXT sensors and actuators are heterogeneous and return different numbers and types of outputs. The device layer constructs a standardised, context-free grammar (CFG) [5] representation of sensor data and available actions and communicates this to the agent layer.

Device Manager: The device manager identifies the sensor devices S_1, S_2, S_3 … attached to the robot. Data from all sensors is encapsulated as a single sensation **S** represented as a variable length string of label-value pairs. The device manager also identifies the actuators A_1, A_2, A_3 … attached to the robot.

Resource Manager: The resource manager communicates sensations and actions between the agent and the devices via Bluetooth. In the demonstration in Section 3, the agent layer is run on a PC separate from the NXT brick. The Lejos Java firmware does not currently offer sufficient support for the complex programming constructs required to implement a curious agent. In future it is envisaged that it will be possible to run the agent layer on the NXT brick and the resource manager will no longer be required.

2.2 Agent Layer

The agent layer comprises a motivated reinforcement learning agent [5] using a computational model of curiosity as the motivation function. The agent has four processes for perception, curiosity, learning and activation. These communicate with the physical sensors and actuators via the device layer.

Abstract Sensor: The abstract sensor communicates with the resource manager via Bluetooth and receives a sensation **S** and set **A** of available actions at each time-step t.

Perception: The perception process computes an event representing the change between the current sensation and the previous sensation stored in memory. Events are also represented as variable length strings of label-value pairs. The value of each event element is computed as the real number difference of sensation elements with the same label.

Curiosity: The curiosity process computes a curiosity value for the current event. Curiosity is computed as a function of the novelty of an event. Novelty is calculated using an Habituated Self-Organising Map [4] such that novelty depends on the similarity of the current event to previously experienced events. The curiosity value is computed using the Wundt [12] curve, shown in Figure 2, so that the most curious events are those that are moderately novel in the agent's experience.

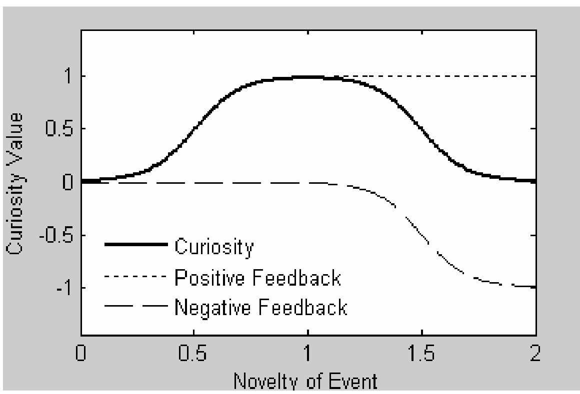


Fig. 2. Model of curiosity using the Wundt [12] curve

Learning: The learning process uses Q-learning [10] to update a state-action table stored in memory. This table maps sensations to actions and utility values. Utility is computed using the curiosity value as the reward in the Q-learning update. Because the agent will compute high curiosity for different events at different times based on its experiences with its current sensors and actuators, the agent will learn different behaviours that are adapted to its current sensors and actuators and current structure.

Activation: The activation process uses an ε-greedy algorithm to select the action with the highest utility value from the state-action table 90% of the time and a random action 10% of the time. This means that the agent follows its learned behaviours 90% of the time and experiments 10% of the time. This encourages the robot to adapt.

Abstract Actuator: The abstract actuator communicates the unique identifier of the action selected by the activation process to the resource manager.

3 Demonstration of a Curious Robot for Creative Play

This section describes a proof-of-concept demonstration of a curious, reconfigurable robot. A designer plays with the robot and changes its structure. The robot adapts its behaviour as different sensors and actuators are added, inspiring the designer to further modify the structure. The designer builds a jointed limb as shown in Figure 3. Figure 2(a) shows the lower part of a limb, mounted on a turn-table with an attached compass sensor. In this structure the curious robot learns behaviours to rotate the limb, as the agent layer is curious about changes in the compass reading.

The designer then develops the limb further into a jointed arm, as shown in Figure 2(b). The arm can extend and retract, has a touch sensor as a finger and an accelerometer to measure its tilt. In this structure, the robot is curious about changes in the accelerometer reading and learns behaviours to extend and retract the arm.

The designer then restructures the limb into a jointed neck, as shown in Figure 2(c). The neck can extend and retract and has an ultrasonic distance sensor. In this structure, the robot is curious about extension and retraction of the neck, which cause changes in the distance readings. The robot also develops a somewhat creative behaviour in which it rapidly extends and retracts the neck. The resultant jerking of the structure causes changes in the compass sensor reading, which the agent finds curious.

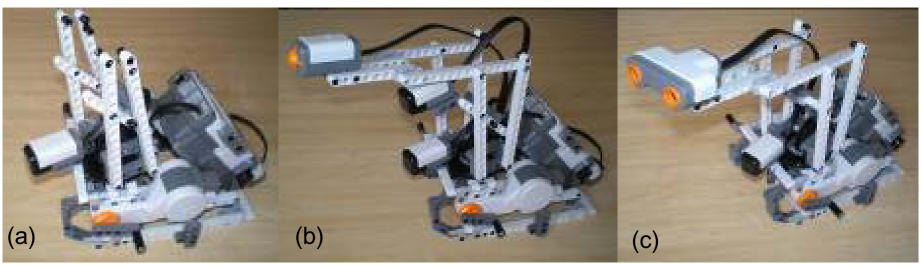


Fig. 3. Reconfiguring a curious robot limb. (a) Lower limb and compass sensor (b) Jointed arm with touch sensor and accelerometer (c) Jointed neck with ultrasonic distance sensor.

The examples above show how the curious agent learn behaviours for different structures, without requiring changes to the agent model. This encourages the designer to play with relationship between structure and behaviour. However, a number of issues arising from the combination of the curious agent with the reconfigurable robot were also identified in this pilot study. First, high resolution sensors can make learning of behaviours quite slow and introduce noise that inhibits learning. In this study, for example, the resolution of the compass sensor was reduced from 360 degrees to 8 compass points to speed learning.

The second issue identified is that the physical structure of the robot may include limitations that are not programmed for in the generic curious agent. For example, the extension and retraction of the limb in this study are bounded by the maximum and minimum angle that can be achieved by the configuration of vertical and horizontal beams. However the servo motors are relatively strong, so the curious agent can experiment with the limb to destruction by forcing it past its limits. This suggests that future reconfigurable robots will have to utilise more flexible joint structures. For example, cog setups where continuous forward (or backward) motion of a motor results in a cycle of retraction and extension.

4 Conclusion

This paper has presented an architecture for curious, reconfigurable robots for creative play. Curious robots can learn new behaviour in response to changes in their structure. The robots in this paper use a computational model of curiosity to identify new goals and reinforcement learning to develop new behaviour in response to those goals. An initial demonstration of curious robots using the Lego Mindstorms platform shows how curious, reconfigurable robots can influence the creative thinking spiral. The adaptive behaviour of the robot in response to its changing structure encourages reflection, imagination and ongoing creative play.

In future, this research may proceed in two directions. First, user studies can provide insight into the impact of this technology on creative play. In conjunction with this, the technology itself can be further developed with the aim of designing robots that can develop complex, realistic and meaningful behaviours. Such robots might have applications beyond entertainment robots to explorer robots in spaces or assistant robots in the home or industry.

References

1. Boden, M.: The creative mind, myths and mechanisms. Wiedenfield/Nicholson, London (1991)
2. Gero, J.S.: Creativity, emergence and evolution in design. In: Second International Roundtable Conference on Computational Models of Creative Design, Australia, pp. 1–28 (1992)
3. Kim, S.H.: Essence of creativity. Oxford University Press, New York (1990)
4. Marsland, S., Nehmzow, U., Shapiro, J.: A real-time novelty detector for a mobile robot. In: EUREL European Advanced Robotics Systems Masterclass and Conference (2000)
5. Merrick, K., Maher, M.-L.: Motivated reinforcement learning for adaptive characters in open-ended simulation games. In: ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE 2007), Salzburg, Austria, pp. 127–134 (2007)
6. Oudeyer, P.-Y., Kaplan, F.: Intelligent adaptive curiosity: a source of self-development. In: Proceedings of the fourth international workshop on epigenetic robotics, pp. 127–130 (2004)
7. Resnick, M.: Sowing the seeds for a more creative society, Learning and Leading with Technology. International Society for Technology in Education, pp. 18–22 (2007)
8. Russel, S., Norvig, P.: Artificial intelligence a modern approach. Prentice-Hall, Englewood Cliffs (1995)
9. Schmidhuber, J.: Curious model building control systems. In: International Joint Conference on Artificial Neural Networks, pp. 1458–1463. IEEE, Singapore
10. Sternberg, R.: The nature of creativity. Cambridge University Press, Cambridge (1988)
11. Watkins, C., Dayan, P.: Q-learning. Machine Learning 8(3), 270–292 (1992)
12. Wundt, W.: Principles of physiological psychology. Macmillan, New York (1910)
13. Yim, M., Duff, D., Roufas, K.: Polybot: a modular, reconfigurable robot. In: Proceedings of the 2000 IEEE International Conference on Robotics and Automation, CA, pp. 514–520 (2000)

Musical B-boying: A Wearable Musical Instrument by Dancing

Minoru Fujimoto[1], Naotaka Fujita[1], Yoshinari Takegawa[2], Tsutomu Terada[1], and Masahiko Tsukamoto[1]

[1] Graduate School of Engineering, Kobe University
[2] Organization of Advanced Science and Technology, Kobe University
{minoru,nfujita}@stu.kobe-u.ac.jp, {take,tsutomu}@eedept.kobe-u.ac.jp, tuka@kobe-u.ac.jp

Abstract. Advances in computer technologies have enabled new means of musical expression with motion enabled musical instruments, which have attracted a great deal of attention. We created a new performance system that enables a dancer to make music while b-boying, which is a dance style. We implemented sensor-equipped shoes and a system that recognizes motions and that controls music. We used an implemented prototype on stage, and we found that dancers were able to control the music flexibly while b-boying. Our system will bring about a new style of musical performance for both musicians and dancers.

1 Introduction

Advances in computer technologies have enabled new means of musical expression with motion enabled musical instruments, which have attracted a great deal of attention[1][2][3][4]. These instruments enable users to generate and control the sound by physical motion, and they have led to the development of a new style of entertainment that has the characteristics of integration between physical performance and musical performance. Although a lot of research and art performances have such characteristics, they does not fulfill the requirement that players want to keep the procedure of whole performance under their control. We created a new dance performance system that enables dancers to generate music while dancing. We implemented sensor-equipped shoes and a music control system with these shoes. Our system not only generates sounds when people dance, but it also controls the performance scenario by changing the set of sounds generated by motion commands. Using our system, dancers can flexibly control sound generation while dancing, thereby enabling a new style of musical performance for both musicians and dancers.

2 B-BOYING

B-boying is commonly known as "*breakdancing*", which is a type of street dance. It consists of four elements: entry, footwork, power moves, and freeze. Entry refers to upright dancing. Footwork is a dance performed on the floor. A power

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 155–160, 2008.

move is the most impressive performance because it includes acrobatic moves. A freeze is a stylish pose in reaction to certain beats and is also used to end a performance.

Combining these techniques allows for an infinite number of expressions. Moreover, the dancer adds soul to his/her routine by using various dancing motions including delicate hand-movements and dynamic movements using the whole body.

In addition to the dancing techniques, B-boying has two types of performance styles: showcases and dance battles. In showcase style, people dance to the music. However, in a dance battle, they face each other and compete in turns with improvised performances. Each player has to change his/her expression dynamically in response to the music and the movement of opponents.

3 System Design

As stated previously, b-boying has an infinite range of expression. Additionally, if the dancers were able to generate/control sound with their dancing, a new style of performance would emerge. An integration of dance and sound creation would allow dancers to control everything on stage, and the quality of their performance would improve. Moreover, dancers in dance battles would need dancing and musical skills, raising the interest in dance battles. Therefore, we propose a new style of performance where dancers control not only their motions but also sound generation. In addition to this, if the system has a function to change the allocation between motions and sounds by the dancer's actions, they can fully control the performance on their own.

3.1 Requirements

In usual dancing performances, we dance to the music passively. Dancers or choreographers statically compose the choreography considering the flow and tempo of the music. The configuration of dancing is static, and changing the configuration in response to problems and atmosphere is difficult.

Our system enables dancers to control their performance dynamically by changing the sound generated while dancing. Our pilot study enabled us to clarify these system requirements:

No restriction. The equipment must be small and light enough so as not to restrict movement to enable the quality of dance to remain high. In particular, the devices must be wireless.

Motion flexibility. To play/control music with movements, the system should allow users to configure the relationship between a dance step and a phrase flexibly. Moreover, various types of motions should be accepted by the system.

Personalization. There are various styles of dancing, and motions vary among individuals. Therefore, the motion detection mechanism should be personalized for each user.

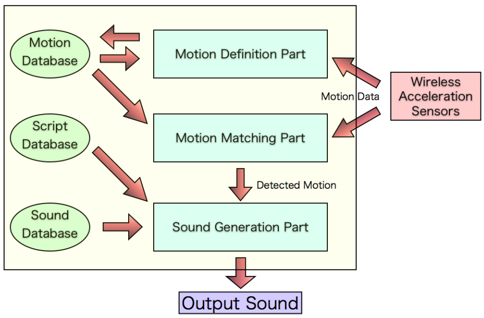


Fig. 1. System structure

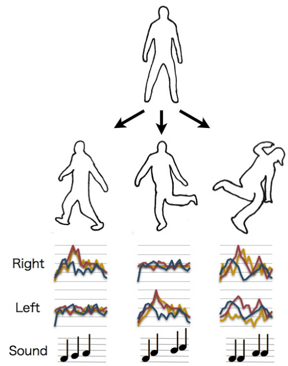


Fig. 2. Example of motion detection

Configuration flexibility. A dance showcase is a kind of story. Therefore, the system should support a change in setting. The system should enable dancers to change a set of motions and allocated sounds according to the change in the setting.

3.2 System Structure

Figure 1 shows the system structure of our system. It consists of 3-axis wireless acceleration sensors equipped in both shoes and a PC application for processing acquired sensor data and for generating sounds. Because dancers only have to wear small sensors on their shoes, the system does not interfere with their dancing (*No restriction*). Our system uses feet motions to recognize dancers' actions. This is because b-boying has many different characteristics in the way dancers move their feet.

The flow of our system is as follows:

Step0. Before using the system on stage, the user registers motions and scripts.

Step1. While a performer dances, the motion of his/her feet is continuously transmitted to the system on a PC.

Step2. The system recognizes the motion by comparing the acquired motion and stored motion data from a *motion database*.

Step3. By retrieving actions allocated to the recognized motion using a *script database*, the system decides the action to take such as generate sounds or change the set of commands/sounds.

When dancers first use our system, they register their motions in it. To register a motion, a dancer performs the actual motion while wearing the acceleration sensors on his/her shoes. This creates g*motion flexibility*h and g*personalization*h because the user can register any motion freely. Then, the dancer describes the scripts. A script includes various definitions of mappings between motions

Table 1. Commands used in script

Name	Function
right *act*	recognition of *act* on right foot
left *act*	recognition of *act* on left foot
both *act*	recognition of *act* on both feet
set *grp*	creation of a command group *grp*
play *snd*	specification of the sound *snd* to play
change to *grp*	jump to a command group *grp*
for n times	n times recognition is required
after *act*	sequential recognition after *act*

set *block1* ...(1)
right *Sample1* play *Music1* ...(2)
left *Sample1* change to *block2* ...(3)
set *block2*
both *Sample1* for 3 times play *Music2*...(4)
left *Sample1* after left *Sample2* play *Music3*...(5)
right *Sample1* change to *block1*

Fig. 3. Example of script description

and commands, such as generating sounds, changes in music, and changes in a mapping set (the detailed functions in a script are described in the next section). It contributes to the *Configuration flexibility* of the system.

When used on stage, the wearing sensors continuously transmit the motion data to the PC application via wireless communication. The system recognizes the motion by comparing the acquired motion and stored motion data from the *motion database*. Our system uses a dynamic programming (DP) matching algorithm to analyze the acceleration-sensor data. DP matching is a well-known method for retrieving similarities in time series data such as speech or motions. Because each sensor outputs 3-dimensional acceleration data, the system acquires 6-dimensional data for comparison.

The system retrieves the actions allocated to the detected motion using the *script database*. For example, Figure 2 shows the generation of three types of sounds according to three different motions.

3.3 Script for Dancing

To achieve flexibility, we use a scripting mechanism for enabling dancers to change the system settings as they like. Our script consists of a set of commands, which are shown in Table 1.

Figure 3 shows an example of a script description. Each part of the figure means the following functions:

1. Definition of a command group *block1*: this group continues until another group definition.

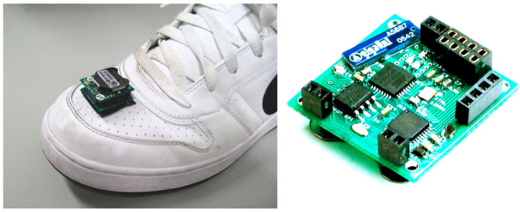


Fig. 4. Acceleration sensor with shoes

2. The system outputs *Music1* when *Sample1* motion is recognized on the right foot.
3. The system changes the group to *block2* when *Sample1* motion is recognized on the left foot.
4. The system outputs *Music2* when *Sample1* motion is recognized on both feet three times in succession.
5. The system outputs *Music3* when *Sample1* motion is recognized on the left foot sequentially after the recognition of *Sample2* on the left foot.

These functions enable a dancer to utilize the system flexibly. In particular, the group definition function contributes to the flexibility and accuracy of recognition. This function enables the system to allocate different sounds to one motion, and because unnecessary motions are excluded for recognition, the recognition accuracy increases compared with a case where all motions are recognized.

An example of using the grouping function on an actual stage is a dancer playing drum-music in the first half of the show, then playing bass-music in the second half.

4 Implementation

4.1 Hardware Implementation

As described in Section 3.1, we needed small wireless 3-dimensional acceleration sensors on dancing shoes. No suitable sensor was available in commercial products, so our research group developed a new wireless sensor module named the "Nao-RF Chip". This module has a 70Hz 3-dimensional acceleration sensor, and the size and weight allow it to be attached to shoes.

Figure 4 shows the sensor module and a shoe with the module.

4.2 Performance

We used our system in several actual dance performances. One example of such performances was a showcase at the *Kobe Luminarie Live Stage* on December 8th and 9th, 2007. Kobe Luminarie is a light festival held in Kobe, Japan every December. It began in 1995 and commemorates the Great Hanshin earthquake of that year.

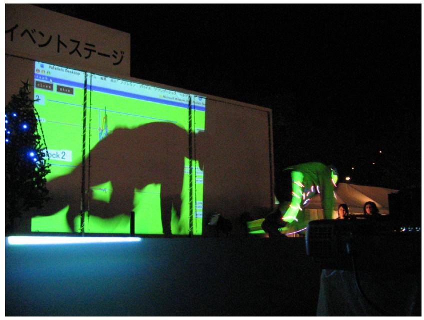


Fig. 5. Snapshot in Kobe Luminarie (1) **Fig. 6.** Snapshot in Kobe Luminarie (2)

Our show consisted of various types of dancing, as shown in Figure 5, 6. In the show, two dancers improvised performances in turn. Using our system, they changed the music pattern while dancing. Moreover, because we added a function to change the projected background color in response to motion, the audience became more interested in the performance.

5 Conclusions

We described a new style of system that enables dancers to make music while they dance. Our system uses a scripting mechanism to control performances flexibly. We focused on a person-centered design where users can customize everything, and this characteristic contributes to enabling their emotions to be expressed intuitively and to creating a new type of entertainment. Actual use of the system on stage revealed that dancers could effectively make music.

In the future, we will create an algorithm with high recognition accuracy. In addition, we will apply our system to various types of dancing styles other than b-boying.

References

1. Aylward, R., Paradiso, J.: Sensemble: A Wireless, Compact, Multi-User Sensor System for Interactive Dance. In: Proc. of the International Conference on New Interface for Musical Expression, pp. 134–139 (2006)
2. Bromwich, M., Wilson, J.: 'BODYCODER' A sensor suit and vocal performance mechanism for real-time performance. In: Proc. of the International Computer Music Conference, pp. 292–295 (1998)
3. Tanaka, A.: Musical Technical Issues in Using Interactive Instrument Technology with Application to the BioMuse. In: Proc. of the International Computer Music Conference, pp. 124–126 (1993)
4. YAMAHA MIBURI (1990), http://www.yamaha.co.jp/design/products/1990/miburi/

Hybrid Visual Tracking for Augmented Books

Hyun S. Yang[1], Kyusung Cho[1], Jaemin Soh[1], Jinki Jung[1], and Junseok Lee[2]

[1] Department of Computer Science, Korea Advanced Institute of Science and Technology, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea
[2] Electronics and Telecommunications Research Institute, 138 Gajeongno, Yuseong-gu, Daejeon 305-700, Republic of Korea
{hsyang,qtboy,jmsoh,jk}@paradise.kaist.ac.kr, leejs@etri.re.kr

Abstract. The augmented book is the system augmenting multimedia elements onto a book to bring additional education effects or amusement. A book includes many pages and many duplicated designs so that tracking a book is quite difficult. For the augmented book, we propose the hybrid visual tracking which merges the merits of two traditional approaches: fiducial marker tracking and markerless tracking. The new method does not cause visual discomfort and can stabilizes camera pose estimation in real-time.

Keywords: hybrid visual tracking, augmented reality, augmented book.

1 Introduction

Recently, there have been a variety of approaches to raise educational achievement as well as enjoyment of books. As an example of these approaches, some systems have been developed to bring additional education effect or amusement in the augmented reality field, augmenting multimedia elements onto a book. Billinghust et al.'s *Magic Book* [1], Taketa et al.'s *Virtual Pop-up book* [2], and Cho et al.'s *e-Learning system* [3] are good examples. We will call this kind of systems augmented books.

Like other augmented reality systems, the most important problem of the augmented book is the registration between the real and virtual worlds. To address the registration problem, fiducial marker tracking or markerless tracking approaches are usually used. Fiducial marker tracking can support enough IDs and is fast, but it causes visual discomfort. On the other hand, markerless tracking does not cause visual discomfort and some real-time methods were proposed, but using this approach for tracking pages of a book is not adequate because a book includes tens or hundreds of pages and many pages are very similar, making it difficult to distinguish each page and, moreover, handle in real-time.

Applying either approach to the augmented book brings pros and cons, and pros and cons of either approach are opposite to the other one. We propose a hybrid method which can overcome the shortcomings of both approaches. That is, the new approach does not cause visual discomfort and can stabilize the camera pose estimation in real-time.

The remainder of this paper is organized as follows. In Section 2, we introduce fiducial marker tracking and markerless tracking methods and analyze their merits and

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 161–166, 2008.

demerits. Section 3 focuses on the hybrid visual tracking, while Section 4 explains its implementation. Section 5 presents the results and Section 6 concludes this paper.

2 Related Work

In fiducial marker tracking, a fiducial marker is surrounded by a black rectangle or circle shape boundary for easy detection. ARToolkit [4], Matrix [5], and Cantag [6] are representative examples of fiducial markers. For the rectangular shape case, the camera pose is estimated using the projective relation between vertices in the scene and the world. A larger marker size is better for accurate and stable camera pose estimation, but users feel visually uncomfortable with large marker. Visual comfort and accuracy of camera pose estimation are in a tradeoff relation. In addition, almost all fiducial markers include a bit pattern representing IDs which are detected quickly and easily with binarization of the whole scene.

In markerless tracking, the key issue is the keypoint matching between scenes taken from different viewpoints. Many keypoints spread in the whole scene so that if keypoint matching is accurate, a much more accurate camera pose can be obtained compared to fiducial marker tracking. Furthermore markerless tracking does not cause visual discomfort because it does not use any fiducial markers. Famous algorithms include Mikolajczyk et al.'s method [8], SIFT [9], and SURF [10]; however these methods consume a great deal of time to create descriptors, so they are not suitable for augmented reality systems, which require real-time performance.

Recently, Lepetit el al.'s keypoint matching method [7] using randomized trees have been fast enough to perform in real-time and has been robust to various viewpoints so that it come into the spotlight of the community. Lepetit el al. transform a local image patch of each keypoint into the almost possible appearances, and train N randomized trees with those transformed patches. Williams et al. [11] suggest the modified randomized trees, which make the real-time training possible, and apply this method to the real-time SLAM. These two methods are state of the art real-time keypoint matching. However, these outstanding works are not suitable for the augmented book because they create a very large number of key points for many pages (approximately 10000 keypoints for 100 pages), and require a tremendous amount of memory (approximately 6GB for 100 pages), and consume a great deal of time for matching the large number of keypoints. Furthermore, if some similar pages exist in one book, page identification will be unstable.

3 Hybrid Visual Tracking

As stated in Section 2, fiducial markers can express abundant IDs, and markerless tracking does not cause visual discomfort and is accurate and stable for camera pose estimation. The hybrid visual tracking merges the merits from both approaches. The key idea is that page identification is performed with fiducial marker detection, and camera pose estimation is performed with randomized trees. We will call fiducial marker for page identification as page marker. Fig. 1 shows each page has the trees for keypoint matching. The page marker supports just page identification. After page identification, the system loads the trees for the identified page, performs keypoint matching, and eliminates outliers of a matching result using RANSAC algorithm. If correct matches are enough, the system performs the pose estimation proposed in our previous work [3], otherwise the system detects a page marker because it is likely that the current page is changed.

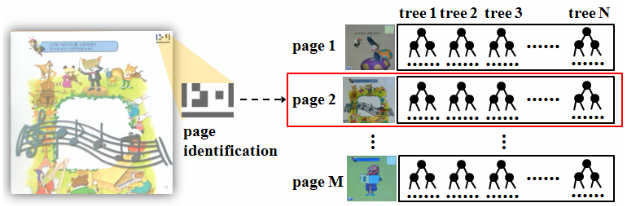


Fig. 1. The concept of Hybrid Visual Tracking

In hybrid visual tracking, a page marker is just used for page identification not for camera pose estimation so that we can reduce the size of a marker and don't need to design the shape as rectangle or circle. Therefore, a page marker can be designed in various ways not giving visual discomfort. Fig. 2 shows that our page marker is very small compared with a traditional fiducial marker. Our marker size is only 2cm x 1.3cm. In addition, randomized trees support only one page and the system loads the trees for only the current page, so we can optimize the memory space and achieve fast matching and high correct matching ratio. Compared to fiducial marker tracking, keypoints spread widely, so more accurate and stable camera pose can be estimated.

Fig. 2. *(left)* a page with a traditional fiducial marker, *(right)* a page with our page marker

4 Implementation

Some requirements are needed for the hybrid visual tracking not to cause visual discomfort and to achieve real-time performance. First of all, the size of page marker is small enough but it must not cause a little false identification. Second, for real-time performance, the keypoint detection must be performed fast enough, and marker detection must not cause overhead. Finally, an efficient data structure for randomized trees is considered to load fast them into memory once a page is changed. We will explain the keypoint detection process, the page marker detection process, and the data structure for randomized trees with considering the above requirements.

4.1 Keypoint Detection

We use the FAST detector [12] which is noticed that it takes 2 ms to find keypoints in a 640 x 480 image in real-time. This algorithm checks whether or not each pixel p is a keypoint. It tests intensities of 16 circular points whose distance is 3 from pixel p. If the absolute intensity differences between p and more than 10 contiguous circular points are all above the threshold value t, p is selected as a keypoint. The magnitude of a keypoint is the sum of the intensity difference between p and all the circular points. This algorithm is very fast because most pixels are rejected by considering only four points instead of all the circular points. We extract 200 keypoints per page for the training step and 150 for the test step using the FAST detector.

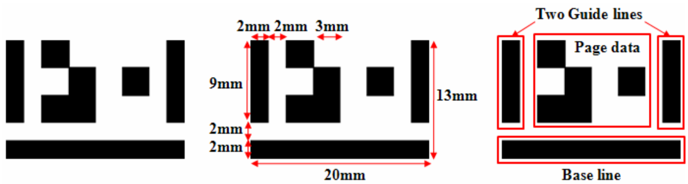


Fig. 3. Page marker design

4.2 Page Marker Design and Detection

For page marker detection, we want to use the result of the keypoint detection to reduce the candidate regions, rather than searching through the entire scene. The FAST detector is sensitive to spiky points and points that have very different intensities from their neighbors. Therefore, we design page markers like in Fig. 3, which includes many spiky points and use just black and white colors. A page marker are surrounded with the base line and the two guide lines, and includes the 12 bits pattern. To reduce the false positive ratio, we divide 12 bits into 8 bits and 4 bits which represent the ID itself and the CRC (Cyclical Redundancy Check) code of the ID, respectively, and also reject the patterns including too many white cells or too many black cells so that only the patterns with between 3 and 9 black cells are used. Finally, the page marker can distinguish 245 IDs without false detection.

For page marker detection, we select the top 50 keypoints based on magnitude and group them in a 30 pixel distance. These groups are candidate regions for a page marker. For each candidate region, CCL algorithm and PCA analysis are performed to find one base line and two guide lines. After detection, page data is decoded and an ID and a CRC code are extracted. If the CRC code is correct, then the ID is recognized as the page ID. This algorithm is very fast because it does not search through the entire scene, but only considers the keypoints detected by the FAST detector.

4.3 The Data Structure for Randomized Trees

When the original randomized trees are loaded into memory from file, tree structures are built one by one, is time-consuming. On the other hand, Williams et al.'s randomized tree can be represented as an array with only d elements instead of a tree with 2^d-1 internal nodes where d is the depth of a tree. By using the latter one, we can make it possible to load randomized trees into memory immediately without building a tree structure.

In practice, we train 40 randomized trees with depth 10 for one page, the total amount of space for one page is 31MB, and it takes 34 ms to load them. As a result, the average correct matching rate is 70%.

5 Results

Our experiments were conducted on a 3.0GHz PC and used a book which has 30 pages. We put page markers only on odd pages and trained randomized trees. In the book, the designs of 5 and 11, 13 and 25, 15 and 27, and 17 and 23 page are exactly the same, which make it very hard to identify the pages with only randomized trees.

The hybrid visual tracking method identified pages very well and estimated camera pose robustly in difficult situations such as viewpoint changes and serious occlusion like Fig. 4. We need only 31MB space of memory to operate the system because we need only one page of data at a time.

Table. 1 shows the average amount of time needed for a 2500 frame video clip for 15 pages. As expected, the page marker detection takes so little time that it does not have an effect on performance. After loading the tree data, only 22.49ms was needed to finish the pose estimation process. That means the proposed method can achieve 44.46 fps ideally.

Fig. 4. Pose estimation under *(left and middle)* viewpoint changes, *(right)* serious occlusion

Table 1. The time needed for hybrid visual tracking

Keypoint Detection	Page Marker Detection	Keypoint Matching	RANSAC	Pose Estimation
6.05 ms	1.11 ms	2.87 ms	6.45 ms	7.12 ms

6 Conclusion

We propose the hybrid visual tracking which can efficiently estimate camera poses when there are many objects to be augmented and many duplicated designs like the augmented book. With only wide-baseline matching for each frame, it takes 22.49ms, but if the small-baseline matching technique and camera tracking techniques such as the extended kalman filter or the particle filter are considered, faster and more robust results can be obtained for camera pose estimation. Furthermore, memory cache and background processing techniques can be used for loading randomized trees to accelerate the page loading speed.

Acknowledgments. This research was supported by the Development of Elemental Technology for Promoting Digital Textbooks and u-Learning, which is sponsored by the Ministry of Information and Communication, and the Ubiquitous Autonomic Computing and Network Project of the 21st Century Frontier R&D Program, which is sponsored by the Ministry of Information and Communication.

References

1. Billinghurst, M., Kato, H., Poupyrev, I.: The MagicBook: A Transitional AR Interface. Computers and Graphics, pp. 745–753 (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. Cho, K.S., Lee, J.H., Lee, J.S., Yang, H.S.: A Realistic e-Learning System based on Mixed Reality. In: 13th International Conference on Virtual Systems and Multimedia (2007)
4. Billinghurst, M., Kato, H.: Collaborative mixed reality. In: 1st International Symposium on Mixed Reality, pp. 261–284 (1999)
5. Rakimoto, J.: Matrix: A realtime object identification and registration method for augmented reality. In: Asia Pacific Computer Human Interaction, pp. 63–68 (1998)
6. Rice, A., Beresford, A., Harle, R.: Cantag: an open source software toolkit for designing and deploying marker-based vision systems. In: 4th IEEE International Conference on Pervasive Computing and Communications (2006)
7. Lepetit, V., Fua, P.: Keypoint Recognition using Randomized trees. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(9), 1465–1479 (2006)
8. Mikolajczyk, K., Schmid, C.: An Affine Invariant Interest Point Detector. In: 5th European Conference on Computer Vision, pp. 414–431 (2002)
9. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 20(2), 91–110 (2004)
10. Bay, H., Tuytelaars, T., VanGool, L.: SURF: Speeded Up Robust Features. In: 9th European Conference on Computer Vision, pp. 404–417 (2006)
11. Williams, B., Klein, G., Reid, I.: Real-Time SLAM Relocalisation. In: 11th IEEE International Conference on Computer Vision (2007)
12. Rosten, E., Drummond, T.: Fusing points and lines for high performance tracking. In: 9th IEEE International Conference on Computer Vision, pp. 1508–1511 (2005)

Towards Emotional Characters in Computer Games

Robert Schaap and Rafael Bidarra

Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
P.O. Box 5031
2600 GA Delft, The Netherlands
R.E.J.Schaap@gmail.com,
R.Bidarra@ewi.tudelft.nl

Abstract. This paper describes our research on integrating emotion aspects into characters of video games. Motivation for this is the significant absence of emotive expression in most characters of current games. In order to provide players with a more immersed game experience, it is essential that characters show compelling expressions, in particular, facial expressions. As a research methodology, we took several intense scenes from classic movies and evaluated what would be necessary to allow a virtual character to reenact the scenes inside a game. The recreated scenes should convey the same emotions as the original scenes. Our approach represents the internal state of a video game characters by using a PAD-based model, as well as modeling the impact on it of many relevant environment aspects. Representing this internal state also involves blending together animations to create lively facial expressions. We conclude that the resulting model is good enough to allow scenes to be recreated with an emotional load similar to the original scenes. The approach is also successful inimproving the consistency of the mood of the characters based on their personality.

1 Introduction

Recently there have been large improvements regarding video games. The advancements in graphics are especially large. Character expression is not getting the attention it deserves to get. Generating interesting and appropriate facial expressions for video game characters is difficult, but it is very important for the overall game-play experience. Most game developers rely heavily on scripting to create the experience they envision. With the expanding scope of current games it becomes more difficult to rely on scripting solely. Game worlds are becoming larger and more complicated. A lot of interaction takes place not just between the players and the game world but also among players and among elements of the game world. Even when we only consider single player games, the game worlds of current games are so vast that it has become very hard to determine when certain events will take place. Even worse, it is often not even possible to anticipate which events will be taking place.

Imagine the situation where the best friend of the player gets shot and right after that something positive happens, so that the latter event would possibly trigger a

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 167–172, 2008.

Fig. 1. The Girl character

script for the player to act happy. This would be very awkward and unrealistic since he just lost a good friend. Of course this could be solved with additional scripting, but this clearly does not scale. The scripts will become complicated and still there will always be situations and sequences of events that the developer did not think about.

With this in mind, we investigated ways to make it possible to create more interesting and emotionally convincing video game characters, as that shown in Fig. 1. To accomplish this, we propose a model that is able to represent the internal state of the characters.

Another goal is to assist the artists who create the virtual worlds and define its inhabitants. Artists do not tend to be very technical people, yet a lot of tools out there are very technical and not very intuitive to use. We want to provide them with easy-to-use tools for defining personality in video game characters. As well as providing them with a straight-forward way of integrating the characters into their environment. To accomplish this we have taken a highly procedural approach which allows for intuitive parameterizations.

In the next section we will introduce a model that is able to represent personality, mood and emotions. We will determine and structure the environmental influences that affect internal state in section 3. Finally, to bring the characters to life, we need to generate actual expressions for the characters based on their internal state. This is detailed in section 4, followed by our conclusions and recommendations in section 5.

2 Representing Internal State

The internal state of a character can be divided into three layers: emotion, mood and personality [1] for which the duration of the effect respectively short, medium and long. See Fig. 2 for a visualization of how the three states relate over time. The Pleasure-Arousal-Dominance (PAD) model is a model that has been developed for convenient assessment of consumer reactions to services and products. This means that the model was meant to describe and record emotions. However, after reviewing the model it

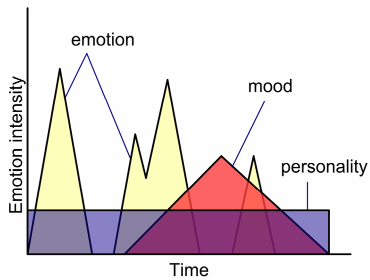


Fig. 2. Emotional intensities over time

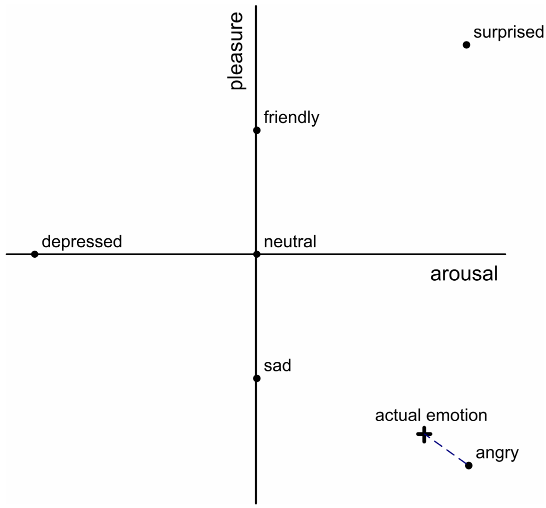


Fig. 3. Emotions in PAD space

became clear that it could also be used for tracking the internal state of characters and provide us with a very good basis for generating facial expressions. We have evaluated the PAD model and compared it to alternative models. In this section we point out the strengths of the PAD model, by which we think it is suitable for use in video games.

The PAD model is based on the view of human emotions as an input-output system [2]. The PAD model lies in between these inputs and outputs. It consists of only three emotional dimensions. These dimensions are Pleasure, Arousal and Dominance. The PAD model has a few properties which make it suitable for representing the internal state of characters. The model has a low complexity due to the fact that it uses only three axes for representing emotions. The PAD model is able to represent a broad range of emotions. It can be compared to creating a whole spectrum of colors using only red, green and blue. The interesting thing about the scales is that they are almost orthogonal to each other. This gives us great flexibility as it allows us to modify the values along the three scales independently without trouble. Fig. 3 shows how a set of emotions can be placed inside the PAD model. For illustration purposes, we left out the dominance axis in this figure, resulting in a crosscut of the actual PAD space.

The pleasure scale signifies the amount of enjoyment or satisfaction the character is experiencing. When the character experiences something good he will feel happy, when something bad happens he will be sad. The arousal scale relates to the amount of arousal or excitement being experienced. For instance, when a character is receiving a lot of information he will be aroused; when there is no information he will be bored. The final scale is the dominance scale which relates to the feeling of being in-control and free. When a character is enslaved, his dominance will be low, when he is free to do what he pleases, his dominance will be high.

Finally, there exist mappings between the PAD model and other well-known models such as Big Five [3]. We use this to provide artists with a set of sliders to tweak personality of characters without having to know about the PAD model.

Another model that is based on the PAD model is ALMA, in which moods are simulated by using a pull-and-push mood change function [4]. We have taken this idea and applied it to our model, so that we can change the current mood based on the active emotion. When the intensity of the active emotion is higher than the intensity of the current mood the current mood gets pulled towards the active emotion. This results in the amplification of the current mood. When the intensity of the active emotion is lower than the intensity of the current mood the current mood gets pushed away from the active emotion. The intensity of the active emotion is used to determine how hard to push or pull the current mood.

After establishing a model for the internal state we created the means to influence this state from the environment by determining the environmental influences.

3 Determining Environmental Influences

Human beings are strongly influenced by their environment. Sometimes they are aware of these influences, but most of the time they do not realize that. Typically, only when these influences become more extreme, people will become aware of them; e.g. extreme heat will certainly be noticed, but a pleasant temperature will not.

Current computer games try to provide gamers with a richer gaming experience. They do this by implementing things like dynamic weather systems and day/night cycles. However, most of the time the characters in these virtual worlds are blissfully unaware of what is happening around them. This makes them seem disconnected from their environment and renders their behavior unrealistic. To fix this we have created a model that takes the effect of the environmental influences into account when determining the internal state of a character.

We started with a survey of influences that are common to characters in computer games. We came to the conclusion we need three basic categories of influences.

These categories are: physical, chemical and psychological influences.

The physical influences come from the body of the character and influence the mood and emotional state of the character directly (e.g. fatigue, illness and hunger). The chemical influences are physical of nature as well, but can be directly linked to substances the character has been exposed to (e.g. alcohol, medicine and drugs). The psychological influences depend on factors that are present in the environment of the character (e.g. freedom, hostility and familiarity). Each of these actual influences has a specific effect on the internal state of the character. For example, fatigue will have a negative effect on the feeling of pleasure and arousal. Alcohol will have a positive effect on pleasure and dominance, but a negative effect on arousal. To accurately model the effects of some influences we needed PAD values that vary over time. Take drugs for example: the way the effects of drugs develop is usually a fast increase right after the drugs have been taken, followed by a slow decay as the drugs are wearing out. New influences can be defined as needed as long as the environmental effects can be expressed in changes through the PAD space.

Now that we have described the internal state and the influences upon it, we have to look at how we are going to make these emotions perceivable to the gamer.

4 Generating Expressions from Internal State

Eyes are a very important part of conveying the internal state of a character. When a character is looking at some items or when he is specifically avoiding them, that says a lot about how the character feels about those items. The way we incorporated the gaze into our model is by relating the arousal scale to the amount of movement of the eyes. A character, who is very aroused, wants to take in as much information as possible and will be looking at many things, frequently switching focus. We also take into account the amount of interest the character has in the objects around him. So he will more frequently and for longer periods of time watch objects that have a higher interest value. On the opposite side, when a character is very non-aroused, he will not look at objects that much. He will stare into empty space, at his feet or at the floor and he will not switch focus very often. The dominance scale will be used for when a character is watching another character. When the value on this scale is high, he or she will look the other character straight in the eyes. When the value is low, he or she will act shy, avoid eye contact and generally look in a downward direction.

Another important part of conveying emotion is the expression on the face of a character. We use the PAD model to determine which emotion will be displayed on the face. In PAD space we have defined a set of emotions (see Fig. 3). For these emotions we have defined actual facial expressions. The actual emotion for a character is marked with a plus-sign. To determine which facial expression should be displayed we take the emotions that are closest to the actual emotion and blend them together using the distance of the actual emotion to the preset emotion as a weight. This means when the actual emotion is not too far from the anger emotion, the character will show a hint of anger. When the actual emotion gets pushed more towards the anger emotion, the intensity of the anger expression on the face will increase. The usual approach in facial animation is to blend together static facial expressions; our approach differs in that we actually blend together facial animations.

Additionally, we add noise to the animations, significantly improving character expression. Even in rest, it is very unnatural for a face of a person to be completely still. The arousal scale of the PAD model is used to determine how much noise to add to the facial animation. Higher values of arousal will result in more and larger movements in the face. This is because a high value for arousal indicates that the character is very excited about something. Blinking is also controlled by the arousal scale. It controls the speed and the frequency of the blinking in much the same way as it controls the added noise. A very aroused person will open his eyes faster and blink more often, whereas eyes that are opening and closing very slow will make a person look very bored or sleepy. All of this makes the characters come better to life.

5 Conclusions and Future Work

The use of the PAD model provides a simple yet efficient way of keeping track of the internal state of video game characters. It is useful for maintaining the mood of characters while taking the personality and the current emotions into account. The categorization of environmental influences gives artists an intuitive parameterization, providing a lot of control over how the environment affects the characters. Generating expressions from PAD values seems to work very well for conveying the mood of a character. The addition of noise based on PAD values definitely makes the characters more interesting and lifelike.

When using our model for conveying really specific emotions, some information is lost when converting these emotions to PAD values and the results are not always very accurate. This is due to the fact that the PAD model consists of only three dimensions, which can result in the loss of some subtleties in emotions. These subtleties could possibly still be brought back with some tweaking or taken into account by extending our model, but this still needs to be researched further. In addition, we will continue to fine-tune and evaluate our model using the prototype we have developed, and to incorporate new ideas with our game industry partners.

For evaluation purposes we have taken some intense classic movie sequences from movies such as “Crouching Tiger Hidden Dragon”, “Pirates of the Caribbean”, “Der Untergang” and “Das Boot”. These scenes have then been replicated using the system we have developed to enhance the animation and to really bring the characters to life. The scenes we take from these movies contain highly expressive characters. One such example is the scene from “Das Boot” in which a submarine crew member completely loses his sanity during an attack on the ship with depth-charges. Due to strong space limitations we refer the reader to [5].

Even though evaluation is still in progress, we can already conclude that the use of the PAD model as the basis for maintaining mood and emotions, as well as for representing them through facial expressions of game characters, is very promising.

Acknowledgements

Thanks to Kim Goossens, Jeroen Krebbers and Igor Frid, of Dutch game development studio Playlogic Game Factory, for watching over the direction of this research project from their game development point of view. We also thank Kim for providing us with the (beautiful) Girl character to work with (Fig. 1). Finally we thank all other people at Playlogic Game Factory for their support.

References

1. Moore, G.: Talking heads: Facial animation in the getaway (2001) (retrieved, February 2006), `http://www.gamasutra.com/features/20010418/moore_01.htm`
2. Mehrabian, A.: The three dimensions of emotional reaction. Psychology Today, 57–61 (1976)
3. John, O.P., Srivastava, S.: The big-five trait taxonomy: History, measurement, and theoretical perspectives. In: Handbook of Personality: Theory and Research. Guilford, New York (1999)
4. Gebhard, P.: ALMA: A Layered Model of Affect. In: AAMAS 2005: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems, pp. 29–36. ACM Press, New York (2005)
5. Schaap, R.: The PAD model as basis for emotion of video game characters, MSc Thesis, Delft University of Technology, Delft, The Netherlands (2008)

Sense Cup: A Design of a New Interactive Holistic Sense Convergence Device for Digital Storytelling

Hyun Sang Cho[1,2], Soohyun Lim[1], Tatsuya Yamazaki[2], and Minsoo Hahn[1]

[1] Digital Media Laboratory, Information and Communications University, 517-10 Dogok-dong, Gangnam-gu, Seoul 135-854, S. Korea
haemosu@icu.ac.kr

[2] Universal City Group, Knowledge Creating Communication Research Center, NICT, 3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0289, Japan

Abstract. In this paper, we propose "Sense cup" as a new design concept of a device for enhancing the digital storytelling by serving interactive five senses convergence experience. We designed the cup for providing holistic sense experiences with beverage that synchronized with digital contents such as movies and music and catch the user reaction to generate interaction between users and contents provider.

Keywords: Sense convergence, Digital storytelling, Immersion, User experience.

1 Introduction

Storytelling is one of traditional sharing methods of knowledge and emotion and digital storytelling is a storytelling enhanced by digital technologies in digital media. The term "digital storytelling" has gotten popularity from the first digital storytelling festival that was held in Colorado at 1995 [10]. Joe Lambert, co-founder of the center for digital storytelling [12] said, "Digital storytelling begins with the notion that in the not too distant future, sharing one's story through multiple medium of imagery, text, voice, sound, music, video and animation will be the principal hobby of the world's people".

The main features of the digital storytelling are flexibility, universality, and interactivity [2]. By the flexibility of digital media, traditional linear (sequential) storytelling can be expanded to the nonlinear storytelling such as hypertext storytelling [3]. The universality helps users generate their own stories easily by using cheap and high performance digital technologies. The success story of YouTube shows the impact of the digital storytelling. The interactivity permits users to affect or change the flow of contents (main and ambient) by their decision at every branch of a story. For interactivity, we expanded the concept not only related with the sequence of a story but also related with holistic sense experience that enhances the connectivity between the content of story and users.

Herbert Marshal McLuhan wrote, "The media is an extension of our five senses as we use each of them to communicate to the rest of the world [7]" in his famous book

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 173–178, 2008.

"The Medium is the Message". He insisted that the five senses are all of the gates to get information from the outer including the environment and the other human beings. We believed that the holistic sense convergence experience would dramatically enhance the interactivity of digital storytelling. However, the current multimedia technologies are overly inclined to the visual and the auditory information because of the technical feasibility. Recently there have been studies on emphasizing holistic sense convergence as experience to promote effective user communication by giving realistic user experience especially for marketing and there are many efforts to embody other senses [1]. The most successful field of sense except visual and auditory is haptic technology for a sense of touch. There are several researches and commercial haptic display products for giving users kinetic feedback and feeling of texture of a virtual object [9]. However, the sense of smell and sense of taste still has limitations and there are few successful examples of a device, which serves five senses convergence experiences. Two main reasons might be the unbalance among each of sense technology level and pursuit of too broad universal solution.

In this paper, we propose "Sense cup" as a new design concept of a device for enhancing the interactive characteristics of digital storytelling by serving interactive five senses convergence experience. We designed the cup for providing holistic sense experiences with beverage that synchronized with digital contents such as movies and music and catch the user reaction to generate interaction between users and contents provider. In the remaining of this paper, we explain the limitation and solution of five senses convergence and introduce the design concept of Sense cup and the several implementation examples as future work.

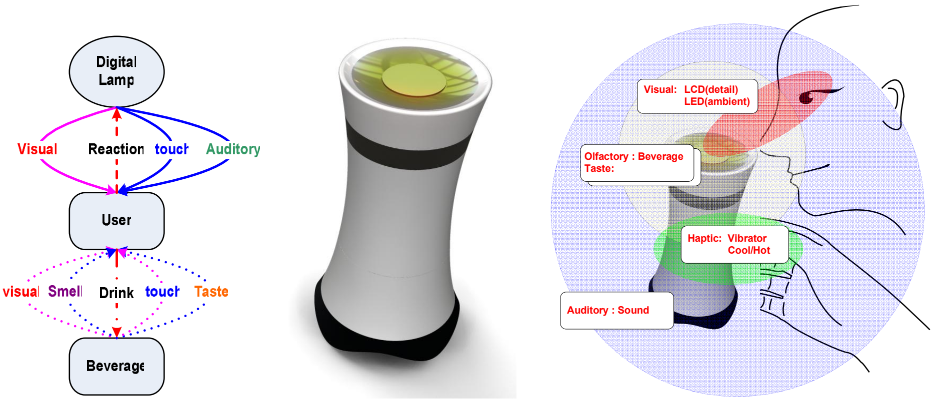
2 Design Concepts of Sense Cup

2.1 Sense Convergence

Sense cup inherited beverage based five senses convergence scenario from our former work "Tea Table Mediator [6]". Fig. 1 (a) shows the convergence relation among five senses. In the former research, senses from tea (or beverage) such as olfactory, taste, haptic, and visual sense are augmented by sense information digital lamp which has visual (LED pattern), audio (sound), and haptic (wind) information display. We shrank the tabletop to a cup.

The most serious problem on holistic sense convergence is providing of a sense of taste and a sense of smell. We will introduce details of our concept of solution at section 2.2. It has five sense display channels (visual, taste, and odor information are provided at a dish on the top of the cup, sound is provided at a bottom-side speaker, and haptic is provided at side of cup) and one physical movement detection channel. The system also designed to detect the movement of cup from a three axes accelometer for deduction of user's status (physical and psychological status such as concentration, boredom, or surprise from inference from the physical movement) which is synchronized with contents. We are going to build a psychological user state model based on hand movement from psychological researches.

One of our sense convergence scenarios is as follow. When a scene of a movie changes to pine tree forest, pine taste beverage with pine odor will be supplied to a



(a) Tea Table Mediator (b) Appearance of Sense cup (c) Range of five senses

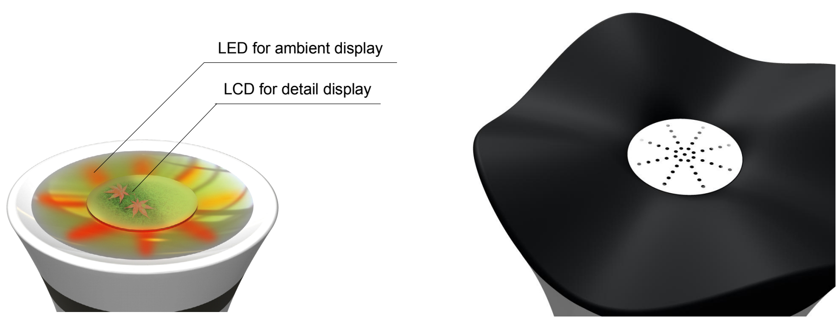
Fig. 1. Sense convergence of sense cup

dish at the top of a cup. In the bottom of the liquid, the LCD shows the detail images of pine trees and the LEDs show waves of tree according to wind. At the same time, the sound and vibration present the movement of wind in the forest.

A cup is one of everyday objects, which serves beverage, and by drinking beverage on morning or at teatime break people can relax and feel, that they are sharing time with others [5]. There have been several trials to enhance the metaphor of cup by digital technologies [4][5].

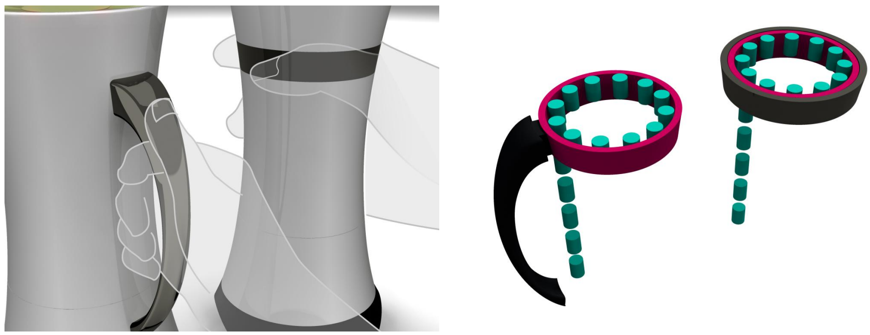
2.2 Displays of Five Senses

The visual display consists of an LCD and a circular LED array. The LCD shows detail information and LEDs show abstractive shape or movement such as ambient of a scene, the waves, and movement of flowers. The detail display sends clear messages for a case of marketing promotion. The ambient display promotes the curiosity for ludic applications [8]. A bottom-side equipped speaker provides sound information.



(a) Visual display at top (b) Sound speaker at bottom

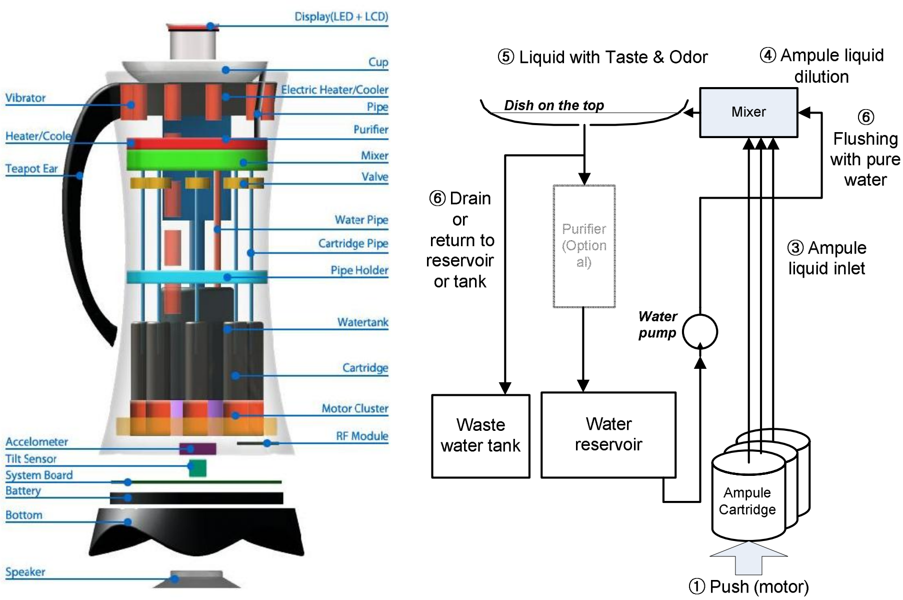
Fig. 2. Visual display and audio system



(a) Hand contact with hapic position (b) Points of vibrators/hot/cool elements

Fig. 3. Haptic display

The haptic display consists of array of vibrators and electrical hot and cool devices. We assume that a position of hands is almost same for every grab (Fig. 3 (a)) and it is possible to serve tactile information by combining the two kinds of information. The vibrators are equipped vertically and circularly. The circular distribution can be expanded to spiral distribution then, for instance, a user can feel the swirly sink into the bottom of lake or whirlwind of a fairy. Electrically operated cool and hot elements are equipped with a thermal conductor band at the location where directly contacts with hands and provide change of temperature instantly.



(a) Internal structure of Sense cup (b) Taste and odor operation

Fig. 4. Taste and odor display

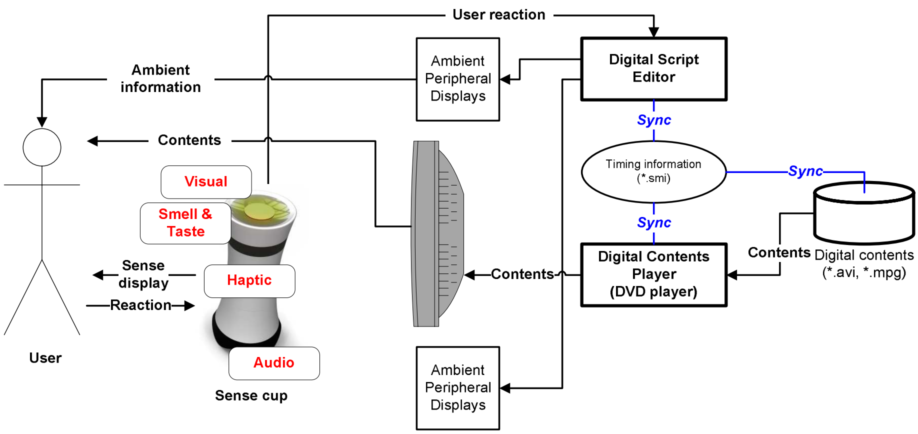


Fig. 5. Configuration of overall Sense cup system

One of the main problems of the sense of taste is a lack of reality of chewy feeling. In earlier stage, a piece of taste-printed paper served a taste. However, it was not successful because of the chewing feeling was mismatched with the taste and the paper was replaced with a wafer [11]. Users could byte and could chew the wafer but it still differed from real feeling. We supposed the solution to overcome the problem by serving a beverage. Since we are familiar with many liquid types of foods such as juice, tea, and alcoholic beverages, it can be possible to solve the mismatch problem. The liquid type food can solve the problem of the sense of smell at same time. The problems of the technologies of odor are; first, there are no definite model and elements of the odor yet, and second, it needs a relatively large volume of air space for serving odors and it is difficult to maintain the intensity and difficult to remove residual odors. In the concept of Sense cup, the odor was generated from the liquid in the dish and the range of the odor is restricted to small space around the dish. A user can smell the odor when he/she wants to drink the liquid or move the cup to near at his/her nose. The odor will be removed quickly as soon as the liquid is drained and flushed. Fig. 4 (b) shows whole process of the taste and odor operation.

3 Overall System Configuration

Fig. 5 shows the whole system configuration. It mainly consists of a digital content play system, a digital script editor, and Sense cup and ambient displays. The digital content play system is conventional devices such as a DVD player or a CD player. The digital script editor edits a timing chart of five senses convergence display script. The timing stamp can be obtained from caption files (for example, *.smi or *.srt files) if the content is a moving picture or it can be generated manually. The sense display information is sent to Sense cup by pre-edited timing sequence through wire (RS-232C/USB/Ehternet) or wireless (Blutooth/Zigbee/WiFi) communication channels. In current design, the sense information in the script is managed separately from content but it can use standardized methods such as Representation of Sensory Effects information (RoSE) of MPEG [13].

As we described in section 2.1, the cup detects movement of user's hands. The movement of hands can show useful information of a user. When a user concentrates on a current content, the movement of a cup is relatively stationary. If a user feels boredom on the content, user will move the cup by various patterns and move the cup abruptly if he/she feels surprise. The editor operates several ambient displays (light/sound/physical movement) for enhancing the user's concentration and interest by the results of the analysis or changes the content as a nonlinear storytelling method.

4 Conclusion

In this paper, we showed the conceptual design of Sense cup as a new device of interactive holistic sense experience for digital storytelling. The digital storytelling has been used in various fields of entertainment, education, marketing, and so forth and keeps expanding its territory. The features of digital storytelling: flexibility, universality, and interactivity can be enhanced by integration of multi-sensory digital media. The current sense convergence has limitation because of unbalance among each of sense technologies. Technologies of a sense of the olfactory and a sense of the taste are relatively insufficient. We proposed Sense cup as a device for digital storytelling and beverage based five senses convergence system as a solution.

References

1. Schmitt, B.H., Simonson, A.: Marketing Aesthetics: The Strategic Management of Branding, Identity and Image. Simon & Schuster Inc. (1997)
2. Choi, H.-S.: Digital Storytelling. Communications of the Korea Information Science Society 21(2) (2003) (Korean)
3. Garzotto, F., Forfori, M.: Hyperstories and Social Interaction in 2D and 3D Edutainment Spaces for Children. In: HT 2006, August 22–25 (2006)
4. Gellersen, H.W., Beigl, M., Krull, H.: The MediaCup: Awareness Technology Embedded in and Everyday Object. In: Gellersen, H.-W. (ed.) HUC 1999. LNCS, vol. 1707. Springer, Heidelberg (1999)
5. Chung, H., Lee, C.-H.J., Selker, T.: Lover's Cups: Drinking Interfaces as New Communication Channels. In: CHI 2006 (April 2006)
6. Cho, H.S., Park, K.S., Hahn, M.: Tea Table Mediator: A Multimodal Ambient Display on the Table Exploiting Five Senses Convergence. In: Ma, L., Rauterberg, M., Nakatsu, R. (eds.) ICEC 2007. LNCS, vol. 4740, pp. 16–21. Springer, Heidelberg (2007)
7. McLuhan, M., Fiore, Q.: The Medium is the Message. The Penguin Press (1967)
8. Gaver, W.W., et al.: The Drift Table: Designing for Ludic Engagement. In: CHI 2004 (April 2004)
9. `http://www.msl.ri.cmu.edu/projects/haptic_consortium`
10. `http://dstory.com/`
11. `http://www.temple.edu/ispr/examples/ex01_01_25a.html`
12. `http://www.storycenter.org`
13. `http://www.chiariglione.org/mpeg/working_documents/explorations/RoSE/RoSE-Reqs.zip`

Frame Selection for Automatic Comic Generation from Game Log

Tomonori Shuda and Ruck Thawonmas

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

Abstract. Recently, we have presented a comic generating system that visualizes an online-game play. Our system was inspired by a former work of Shamir et al. However, comics generated in their work can have series of similar frames when multiple actions occur near each other in both time and space. In this paper, we first present a frame-selection module that uses Habituated Self-Organizing Map. Our method prevents comic readers from boredom by getting rid of resemble consecutive frame candidates. We then evaluate the method by a subject experiment using a play from the ICE, an online-game developed in our laboratory. Experimental result confirms that our method is effective in making output comics more interesting than a baseline method.

1 Introduction

Recently, we have presented a system for generating comics from online-game play log [1]. Figure 1 gives an overview of the system. This system enables users to look back their plays or to introduce their plays to other users. Comic-style outputs have advantages than other-style outputs. For example, the comic-style enables one to distinguish important scenes based on the frame size. And one can grasp the whole story at one glance. In addition, traditional comic techniques can be introduced to achieve expression of various situations. Comic-style representation has been used for summary of conference participation [2], diary experiences [3], and video contents [4], as well as online-game events [5].

Our work in [1] was inspired by the work in [5] and adapted its scene partition method and event extraction method. However, comics generated in [5] can have series of similar frames when multiple actions occur near each other in both time and space. In this paper, we describe and evaluate our solution to this problem, i.e., the frame-selection module in Fig. 1 where Habituated Self-Organizing Map (HSOM) [6] is used. As in [1], we test the system using an online game called the ICE, under development at the authors' laboratory, where typical online-game missions, such as monster fighting and item trading, are available. A screenshot of this game is shown in Fig. 2.

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 179–184, 2008.

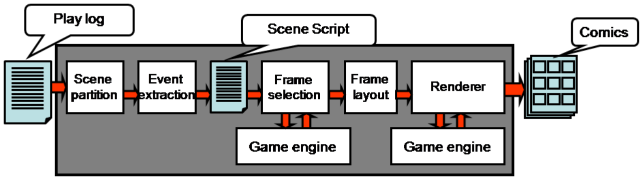


Fig. 1. Overview of the comic-generation system

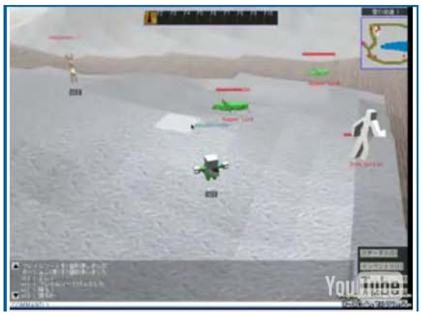


Fig. 2. Screenshot of the ICE

2 Frame-Selection Module

In order to prevent comic readers from boredom, the frame-selection module removes resemble and consecutive frame candidates using HSOM. Frame candidates are generated in the previous module and stored in the scene script (c.f. Fig. 1). Each frame candidate represents an output-comic frame and has frame information such as the frame importance, in the range of 0 to 1, the frame time, and the camera parameters. HSOM enables selection of frame candidates from perspective of comic readers' interest. It is an integration of Self-Organizing Map (SOM) and a habituating layer.

Fig. 3 gives an overview of the HSOM architecture. SOM consists of an array, called the competitive layer, of k competitive neurons arranged in a 2-dimensional rectangular grid. Let assume that there are M input patterns, each corresponding to one of the M frame candidates in a scene of interest, with N dimensions, and that a competitive neuron i has an associated weight vector $\mathbf{w^c}_i = [w^c_{i1}, w^c_{i2}, \ldots, w^c_{iN}]^T$. The habituating layer is an array of k habituating neurons, each connected to the competitive neuron below, having habituating weights w^h_i. The smaller w^h_i means the higher habituating degree. For each scene, frame selection goes through the following steps.

1) Determine the elements of input patterns $\mathbf{x}(t)$, $t = 1, 2, \ldots, M$. These elements consist of the frame time and entities in the entity list. The frame time represents the time when frame candidate t occurs, and it is normalized so that the scene's start time is 0 and the scene's end time is 1. The entity

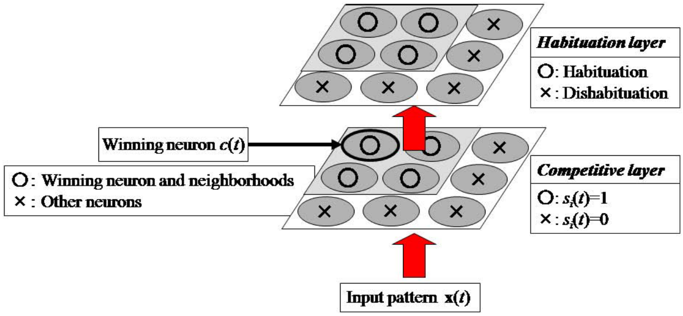


Fig. 3. HSOM architecture

list is the list of all entities in the frame candidates of the scene. If an entity exists in frame candidate t, the corresponding element will be 1; otherwise, it will be 0. Each $\mathbf{x}(t)$ is then normalized, i.e., $\mathbf{x}(t) = \mathbf{x}(t)/ \parallel \mathbf{x}(t) \parallel$.

2) Set the time step $t = 1$ and initialize all competitive-neuron weight vectors $\mathbf{w^c}_i(t)$, $i = 1, 2, \ldots, k$, randomly in the range from 0 to 1, and normalize them, i.e., $\mathbf{w^c}_i(t) = \mathbf{w^c}_i(t)/ \parallel \mathbf{w^c}_i(t) \parallel$. In addition, initialize all habituating-neuron weights $w_i^h(t)$ to 1, which means they are not habituated.

3) Input the input pattern $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_N(t)]^T$ and choose the winning neuron $c(t)$ that is closest to $\mathbf{x}(t)$, i.e., $c(t) = \arg_j \min \parallel \mathbf{x}(t) - \mathbf{w^c}_j(t) \parallel$

4) Select frame candidate t if the sum of $w_c^h(t)$ and its frame importance is more than threshold T_h, and then proceed to 5); otherwise, increment t and return to 3).

5) Update all competitive-neuron weights as follows:

$$\mathbf{w^c}_i(t+1) = \mathbf{w^c}_i(t) + \alpha h_{ci}(t)[\mathbf{x}(t) - \mathbf{w^c}_i(t)] \tag{1}$$

where α is the learning rate, and $h_{ci}(t)$ is the neighborhood function defined as

$$h_{ci}(t) = \begin{cases} 1, & \text{if the competitive neuron } i \text{ is the winning neuron } c(t) \\ & \text{or } c(t)'s \text{ Moore neighborhood.} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Then, normalize $\mathbf{w^c}_i(t+1)$.

6) Update all habituating-neuron weights as follows:

$$w_i^h(t+1) = w_i^h(t) + \frac{\beta(1 - w_i^h(t)) - s_i(t)}{\lambda}. \tag{3}$$

In the above formula, β is the recovery rate. The stimulus value $s_i(t)$ has the same value as $h_{ci}(t)$ of the connected competitive neuron. The constant λ influences both habituation and dishabituation speeds. The values of λ are different for the habituating neuron connected to the winning competitive

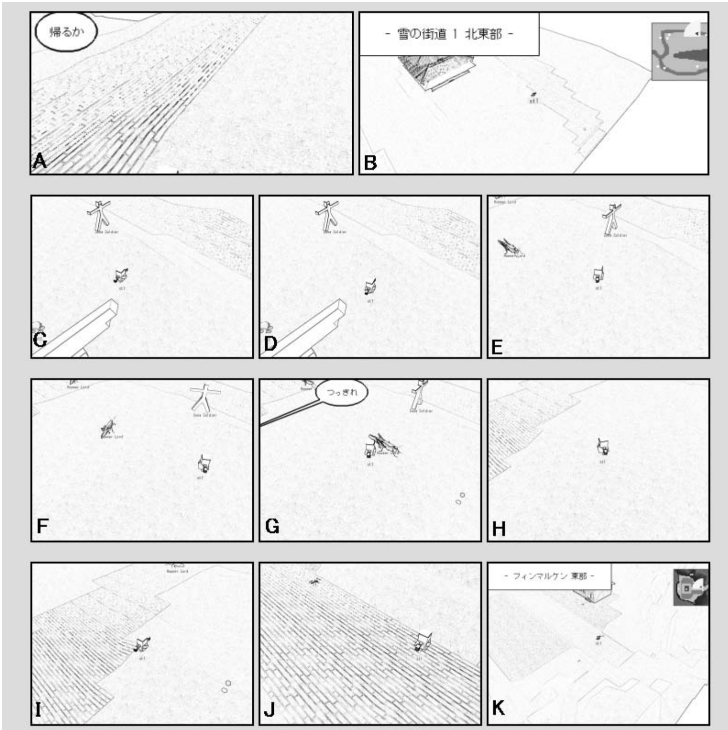


Fig. 4. Example sequence of frame candidates

neuron, for the eight habituating neurons connected to one of the Moore neighborhoods of the winning competitive neuron, and for the remaining habituating neurons.

7) If t is smaller than M, increment t and return to 3).

Figure 4 depicts an example sequence of frame candidates for Comic A described in Sec. 3. They are input to the frame-selection module. Figure 5 depicts the selected frames. Because frame candidates D, F, and I have content similar to C, E, and H, respectively, they are not selected.

3 Evaluation

We tested the frame-selection method with our comic-generation system using the ICE. We requested 12 subjects, who are students in our university, to compare two comics generated from a play[1] of the ICE. These two comics are described below while other conditions, such as camera positions, are identical.

[1] http://www.youtube.com/watch?v=DmvUyl6eZAs

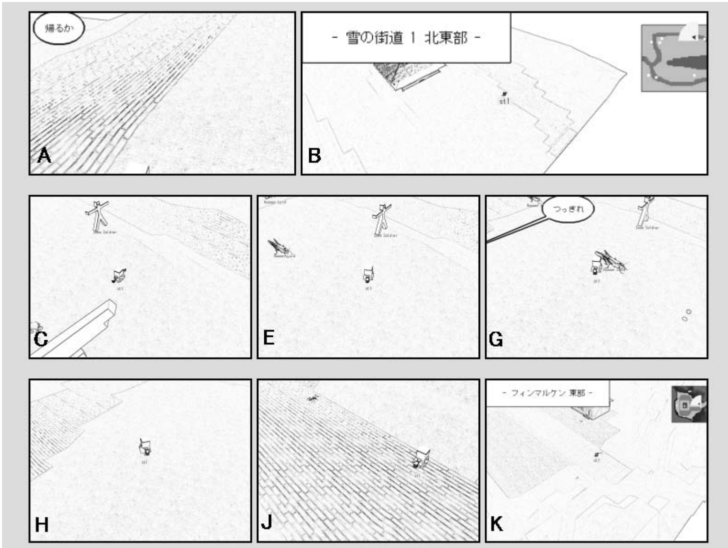


Fig. 5. Selected frames in Fig. 4

– Comic A generated with the proposed frame-selection method described in Sec. 2.
– Comic B generated with a baseline frame-selection method described below using interaction level.

The proposed method selected 52 frames out of 135 frame candidates[2] . The baseline method was also set such that it selected 52 frames; this method selects frame candidates in decreasing order of the smoothed value of the interaction level defined in [1].

We set the number of competitive neurons $k = 16$, arranged in 4 x 4, in order to heuristically limit different situations in a scene to 16. We used $\alpha = 0.25$, $\lambda = 3.33$ for the habituating neuron connected to $c(t)$, $\lambda = 14.33$ for the eight habituating neurons connected to one of the Moore neighborhoods of $c(t)$, $\lambda = 100$ for the remaining habituating neurons. These values were determined following [6]. We chose $T_h = 0.7$ and $\beta = 1.8$ based on our experience because these values were not stated in [6].

We had all subjects watch a video clip of the play in use after explaining them the experiment procedure and how to answer. Then, we distributed the two comics to each subject and had them individually answer which comic fits each of the following questions.

– Which one has the better camera work?
– Which one has the better dynamics?
– Which one represents better the video-clip content?

[2] The number of frame candidates were in the range of 10 to 50 for each scene.

Table 1. Experimental Result

	Camera work	Dynamics	Content representation	Interestingness
Votes for Comic A	9	7	7	7
Votes for Comic B	3	5	5	5
Number of subjects	12	12	12	12

– Which one is more interesting?
– What points of the comics should be improved? (free description)

Table 1 gives the result of the questionnaire except for the fifth descriptive question. It can be seen that Comic A exceeds Comic B for all four questions. This result confirms that our proposed method can reduce resemble camera work and improve dynamics, content representation, as well as interestingness.

4 Conclusions and Future Work

This paper described and evaluated our HSOM-based frame-selection method. The experiment result given in Sec. 3 confirms the effectiveness of this method. Future work includes introduction of a technique that selects frames from story-coherence perspective.

References

1. Thawonmas, R., Shuda, T.: Comic Layout for Automatic Comic Generation from Game Log. In: Proc. 1st IFIP Entertainment Computing Symposium, Milan, Italy (September 2008)
2. Sumi, Y., Sakamoto, R., Nakao, K., Mase, K.: ComicDiary: Representing individual experiences in a comic style. In: Borriello, G., Holmquist, L.E. (eds.) UbiComp 2002. LNCS, vol. 2498, pp. 16–32. Springer, Heidelberg (2002)
3. Cho, S.B., Kim, K.J., Hwang, K.S.: Generating Cartoon-Style Summary of Daily Life with Multimedia Mobile Devices. In: Okuno, H.G., Ali, M. (eds.) IEA/AIE 2007. LNCS (LNAI), vol. 4570, pp. 135–144. Springer, Heidelberg (2007)
4. Calic, J., Gibson, D.P., Campbell, N.W.: Efficient Layout of Comic-like Video Summaries. IEEE Transactions on Circuits and Systems for Video Technology 17(7), 931–936 (2007)
5. Shamir, A., Rubinstein, M., Levinboim, T.: Generating Comics from 3D Interactive Computer Graphics. IEEE Computer Graphics and Applications 26(3), 53–61 (2006)
6. Marsland, S., Nehmzow, U., Shapiro, J.: A real-time novelty detector for a mobile robot. In: EUREL European Advanced Robotics Systems Masterclass and Conference (2000)

Conscientious Objector: Pacifism, Politics and Abusing the Player in Doom 3

Dan Pinchbeck*

Thechineseroom, University of Portsmouth, Eldon Building,
Winston Churchill Avenue, Portsmouth, UK
dan.pinchbeck@port.ac.uk

Abstract. This paper describes Conscientious Objector, a research mod built using the idTech4 engine and assets from the commercial game Doom 3. It was designed to explore the potential for non-lethal force in FPS gaming, and player response to a radically different avatar-NPC relationship.

Keywords: Game, first-person shooter, Doom 3, development-led, modding, pacifism, non-player character.

1 Introduction

Conscientious Objector started life as a simple question: is it possible to build an FPS that removes the ability of the player to actually kill anything whilst retaining the fundamental playability of the game?

FPS games are defined by both the miminal representation of an avatar, and by the essential act of lining up objects with the centre of the screen and delivering an input. This function generally removes the object from the environment. Play thus becomes a process of simplifying the environment until it reaches a critical point, triggering progression. Put more simply, when there is nothing left to kill and no buttons left to press, it's time to move on. All FPS games work to this basic template, whether agents are released en masse in an arcade style, such as Painkiller [1], or there are attempts to model a more `realistic' set of agent-environment relations, like S.T.A.L.K.E.R. [2]. We wanted to attempt to break this normal template and produce an environment that becomes progressively more complex as play progresses. The most obvious way of doing this has the equally interesting side-effect of breaking the other fundamental rule of FPS play, which is killing anything that moves.

Stopping the player from achieving this thus does two things at once: subverts the normal politics of the genre on one hand and establishes a radical new ludic structure on the other.

* Please note that the LNCS Editorial assumes that all authors have used the western naming convention, with given names preceding surnames. This determines the structure of the names in the running heads and the author index.

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 185–189, 2008.

2 Rubber Bullets in UAC Admin

Several pre-existing levels of the commercial game Doom 3 were re-configured for the mod. These were stripped of triggers, agents and objects in preparation for the new gameplay to be levered in. UAC Admin, Mars City and Mars City Underground were chosen as they were small and linear, with the kind of enclosed spaces with limited visibility that would play to the strengths of gameplay designed to engineer as much panic in the player as possible. Without introducing additional items, enough disparity could be introduced into these levels to maintain a sense of progressive gameplay. Additionally, we removed the oxygen from Mars City to introduce a temporal constraint on the player's activities: they were forced to run from oxygen cylinder to cylinder. Using known levels also enabled us, through voice-overs and embedded PDAs, to play a new political dimension off the already existing backdrop of corporate greed and fetishisation of the military.

The damage capacity of the shotgun was reduced to zero, whilst retaining its ability to knock down agents (a simple re-jig of the normal 'getting up' animation found on the basic zombie model). The time this sequence took to run, alongside a reduction in the bounding box and the knock-back effect of a succesful hit, meant that zombies could be effectively blown out of the player's way, giving them time to avoid the agent. Clearly, we are not the first to utilise non-lethal force (NLF) objects in an FPS game: Duke Nukem 3D had both Shrink and Freeze rays [3], Half Life 2's gravity gun occupies a central place in gameplay [4] and, famously, Deus Ex can allegedly be completed using NLF alone [5] – and an early beta of Doom 3 itself featured zombies that could only be permanently killed with a shot to the head. However, in all of these cases, NLF is a supplementary gimmick to core gameplay, rather than a centralised option. Indeed, in the case of Half Life 2, the gravity gun's main use is to hurl distinctly lethal objects like buzzsaw blades, barrels and cars, so it arguably doesn't really count. Likewise, Duke Nukem's ray guns are really there as a means to inflicting a fun coup-de-grace by stamping on or shattering enemies respectively, the latter revived recently by Bioshock [6]. Deus Ex does offer a way for the player to opt for NLF but it may be argued, that this is a niche configuration of gameplay at most. Thus, Conscientious Objector stands as the first FPS system to anchor gameplay directly to NLF. Further, although it may be argued that the mod does not represent a full game, the potential for drawing upon the range of NLF objects already existing in the genre is self-evident, and their addition would easily provide enough diversity of play to maintain the conceit. Indeed, we hope to one day extend Conscientious Objector to the entire of Doom 3 to prove just this. At this stage, however, we will settle for the proof of concept.

3 Politics and Pacifism

We can discuss the politics of FPS games on two levels - the surface and the underlying. In the former case, there is certainly more diversity now than there perhaps used to be. Avatars are still dominantly individualistic libertarians, fighting bureaucracy as much as aliens and demons, and the big businesses and industrial-political complexes and corporations are still corrupt and treacherous, but titles such as Bioshock, S.T.A.L.K.E.R. and

Blacksite [7] are attempting to broaden the political spectrum. Although there is yet to be a return to the more open political and moral choices of the Deus Ex series, Bioshock offers the player the choice of killing or saving the Little Sisters, and Far Cry 2 [8] embeds the consequences of player actions in its world. Likewise, S.T.A.L.K.E.R.'s range of seven endings is based upon political choices the player makes during the game, although to be fair, most of these are not obvious during play. One could also argue that whilst saving the world is nothing new, and revenge following a betrayal by a key non-player character (KNPC) is close to ubiquitous, advances in AI and storytelling have personalised the actions of the player. Half Life 2: Episode Two sees the player fighting to save not millions of lives, but just one; the KNPC Alyx Vance. The very fact that the player is given clear and unambiguous indications that they are supposed to care about Vance and her father is a long way down the road from the faceless, nameless and disposable masses populating earlier games such as Halo (however, it should be noted that alongside saving humanity, the Master Chief is on an equally personal mission to save Cortana in the final installment of the trilogy). Personal motivation, particularly of the protective type, demonstrate a more mature political instinct at work than in early FPS games such as Doom, Quake and Duke Nukem.

However, behind all this increase in sophistication, FPS genres are constructed around a very simple and inflexible political core determined by the central act of gameplay. As John Carmack succinctly put it: “You can dress it up in many ways, but the game still comes down to: go here, touch this, go there, fight...”' [9]. Rather than becoming wrapped up in the politics of violence and the endless discussions of social and psychological impact, the fact remains that FPS games are predicated around the removal of objects from a simplifying environment. Every action in the game is based upon this ludic format. By removing the player's ability to kill - that is, remove - anything, this template is fundamentally shifted, and with it the underlying politics of the game. Once again, it is hoped that with this practical example, we have proved it can be done in a way which critiques the normal surface whilst retaining high focus on gameplay. In other words, it may be a gimmick, but it is a highly playable one.

4 Carl – Abusing the Player

These changes naturally led to the third axis of experimentation in the mod, an abrupt shift to the normal relationship between player and KNPC. There are a number of expected features of this relationship one finds across the genre. Some, like the betrayal of the avatar by the KNPC are functions of extended gameplay, operating as a device enabling both revised interpretation of prior action (thus extending the narrative range of the game) and, in some cases, enabling a virtual increase in complexity without requiring additional assets to be generated. For example, Doyle's betrayal in the last section of Far Cry [10] means that whilst more or less exactly the same course of action is followed by Carver, the emphasis has changed from escape to revenge, the target of the goal has moved from Kreiger to Doyle, and there is a (fake) temporal constraint created by the infection of Carver and Val with the Trigen mutagen. None of this is real, of course, unlike the actual countdown of The Maw, Halo's last level [11] but it does enable players looking for a deeper ludic reality something to get their teeth into. However, regardless of betrayals, KNPCs such as Cortana, or Alyx Vance tend operate as both orientation devices and feedback systems, letting the player know where to go and whether they are acting effectively. This naturally leads towards certain semantic representations, archetypical characters and avatar-KNPC relationships. In Conscientious Objector, we saw an opportunity to subvert these expectations as well. Hence, the game opens with the aggressively cheerful "morning dickhead!"" and ends with Carl, the KNPC triggering the cranial implant that will kill the player as soon as they acheive their goal (he meets the players success and forces what would normally be considered a failure outcome upon them).

Carl operates as the normal political voice of FPS games: reminding the player frequently that they have no power or ability to make decisions, that killing things is both easy and rewarding and pointing out how much more difficult not being able to carry this normal activity out makes the process. Because of his antagonistic position, it also gave us the opportunity to explore pushing the KNPC into as hostile a position as possible. Carl taunts and mocks a successful act as much as a failed one; when the player misses a zombie, they are berated as useless; when they hit one, Carl is quick to point out how little effect it has. In doing so, the frustrations of gameplay find a locus in Carl, the irony being that by the end of play the urge to really, actually shoot someone has been personalised in his direction. Rather than a betrayal creating a forced nemesis, revenge becomes a distinctly localised matter. And, of course, the fact that Carl is openly repellent, coupled with his constant insistence that if only the player acted like an avatar in a normal FPS game would - by killing everything that moved – they would be successful, creates a much more interesting political dynamic in the mod.

5 Conclusion

Conscientious Objector aims to radically break the normal template for FPS play by removing the core act of play: removing objects from the environment. From this deceptively simple start, a range of interesting shifts to expected play naturally fall into line. The underlying politics of gameplay is questioned and thrown into relief, and a new type of player-KNPC relationship is created. But perhaps above all of these experiments, the overriding concern was to create a mod that both explores research questions in a way that simple theorising cannot, and remains a highly playable experience. That is to say, it also stands as an argument that gameplay and research are wholly compatible.

Acknowledgments

Conscientious Objector was built as part of thechineseroom, a development-led research project funded by the Arts&Humanities Research Council, UK.

References

1. People Can Fly: Painkiller. Activision, PC (2005)
2. GSC Game World: S.T.A.L.K.E.R.: Shadow of Chernobyl THQ / GSC, PC (2007)
3. GT Interactive: Duke Nukem 3D 3D Realms, PC (1996)

4. Valve: Half Life 2 Valve, PC (2004)
5. Ion Storm: Deus Ex Eidos Interactive, PC (2007)
6. 2K Games Bioshock 2K Games, PC (2007)
7. Midway: Blacksite: Area 51 Midway, PC (2007)
8. Ubisoft Montreal: Far Cry 2 Ubisoft, PC (2008)
9. Kushner, D.: Masters of Doom. Piatkus (2003)
10. Crytek: Far Cry Ubisoft, PC (2000)
11. Bungie: Halo: Combat Evolved Microsoft Game Studios, XBox (2007)

Experiences Employing Novice Wizard Operators in a Gallery Setting

Steven Dow and Blair MacIntyre

School of Interactive Computing and GVU Center
85 TSRB, 5th St, NW, Atlanta, GA 30308
{steven,blair}@cc.gatech.edu

Abstract. During a eleven-week installation of an immersive augmented reality (AR) experience at an art and technology gallery, we faced the practical challenges of preparing for live public audiences and of training and motivating a group of nine modestly-paid undergraduate museum docents to not only usher the show, but to perform as "wizards". The docents played an integral part of the AR Façade experience, replacing imperfect recognition technology to achieve speech and gesture interaction for the players. Our investigation of the wizard docents revealed insights into two conceptually different wizard-of-oz implementations. The docents initially resisted the interface option that required more cognitive load, but after learning how to make it work, many preferred that interface because it provided them more agency within the system and freedom to diversify the experience for audiences.

Keywords: Immersive entertainment, gallery installations, user experience, wizard-of-oz methods, augmented reality.

1 Introduction

Interactive entertainment experiences—especially in amusement parks or gallery settings—often employ full-time operators to usher visitors in and out of the experiences [5][7]. More recently, interactive media have looked to trained actors performing behind the scenes in real-time as animated characters or robots, such as [8] and [9], very similar to the hidden wizard in the movie *The Wizard of Oz* [1]. Other experiences use actors more overtly, such as *Can You See Me Now?* where online players try to locate physical actors who are moving around a city [3]. In this paper we report on our experience creating an intermediate role for "wizard" docents, who hold responsibilities somewhere between an usher and an actor.

Our discussion centers on an eleven-week public gallery installation of AR Façade, a first-person immersive augmented reality (AR) experience where a player enters a physical apartment with a head-mounted display and interacts using speech and gesture with two fictitious characters [4]. Rather than use speech/gesture recognition technology—which can be problematic, particularly with diverse users in non-controlled settings—we tasked our wizard docents with accomplishing this interaction.

In this paper, we describe the modifications we made to the AR Façade experience to prepare for a long-term deployment at the Beall Center for Art and Technology at

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 190–196, 2008.

the University of California Irvine [2]. The gallery installation would be held to higher standards than our initial lab prototype and it would need to entertain audiences of people, not just the individuals who would participate in the immersive aspect. More importantly, the show would have to be easy to operate by non-technical, undergraduate, gallery docents who were not part of the research team, nor particularly motivated. We set out to investigate whether and how docents would fulfill this role by loosening the bounds of the wizards' task and establishing two conceptually different wizard interfaces. We share data about how the nine docents learned to use and adopt the two interfaces over the course of the show. We found that the interface that required more cognitive processing actually encouraged deeper involvement in the story and in reading the tone of the players and audience members.

2 The Gallery Installation of AR Façade

In this section, we describe our design choices for the three-month Beall Center deployment of AR Façade. We believe our experience of upgrading the prototype into a real deployment can be useful for producers and designers of immersive experiences. We built the immersive interface to be robust and to withstand eleven-weeks of public use by audiences of all shapes and sizes and ages. We designed the physical layout so that audiences and the wizards would get a strong sense of what the player experienced. Finally, we experimented with the wizard infrastructure, providing two methods of asserting player inputs so that we could investigate wizard behavior.

2.1 Physical Setup

AR Façade would be part of a free-to-the-public art gallery for 11 weeks, so we devoted a fair amount of time to the construction of the head-mounted display (HMD) (Fig. 1 *left*) and the physical layout of the space (fig. 1 *middle*). We constructed the head-mounted display from third-party components and custom-created mounts to be robust enough to withstand mishandling. We tracked the HMD using the IS-1200 Vistracker and large visual markers mounted on the 16' tall ceiling. Florescent lights lined the perimeter of the apartment at 10' high and provided an even distribution of light across the ceiling and enough illumination for the stage. We did not encounter

Fig. 1. (left) The head-worn display. (middle) Audience's view into the physical apartment with monitor showing the player's perspective. (right) Player's HMD view shows the Trip character standing behind the physical bar. Note that the black scrim blocks the view of objects beyond the wall.

any noticeable tracking problems—such as jitter and temporary loss of tracking data as we did with our initial prototype—despite loosing a few markers from the ceiling.

We focused on the physical layout of the space, because we wanted the experience to be visually appealing and enjoyable to groups as well as individuals. We used black scrim along one of the long walls of the apartment behind the bar. Since it was light inside the apartment and dark outside, audiences could stand outside the wall and easily see the player's activity, while the player would not be able to see outside though the wall (especially when viewing through the HMD) (fig. 1 *right*). Moreover, the black walls and the black-painted bar helped maintain the occlusion effect when Trip walks behind the bar; players know the bar is there, but the top of the bar is difficult to distinguish from the wall.

We used a wireless video transmitter to transmit the player's HMD view to a TV monitor sitting to the right of the stage. The wizard—sitting just outside the door of the apartment—could also see the same HMD view, along with a second monitor with video from a ceiling mounted camera. The audio turned out to be one of the biggest challenges. We were competing with several other loud exhibits in the space, so we wanted the audience (and wizard) to be able to hear the dialogue between the player and the characters. To make this work, we used two audio transmitters (one to amplify the player's voice and one to transmit the backpack's computer audio) and a mixing board to mix the two sources and disseminate it to speakers for the audience and headphones for the wizard.

2.2 Wizard Infrastructure

We made a number of changes aimed at supporting the undergraduate docents who would be in charge of operating and wizarding the experience. Most of the enhancements were practical (providing a start-up script, building in network checks, etc.), but we also experimented with the underlying infrastructure of Façade. Dissecting the system architecture of the original Façade (described here [6]), we thought it would be interesting to tear out the natural language parser (NLP) and build a wizard interface that could be used to directly trigger the higher-level constructs (discourse acts). This approach was plausible because Façade's AI engine does not explicitly map user inputs to specific statements and actions by the two virtual characters. Instead, it models the characters' emotional states and attempts to choose lines of dialogue based on local and global contexts. There are about 30 possible discourses (e.g. flirt, agree with, etc.) with optional parameters that can be expressed as local context by the player.

The new wizard interface had two panes—"Dialogue" and "Discourses"—both available to the wizard at any time. The Dialogue pane included a text entry field for typing in what players said and large buttons for indicating specific player actions (Fig. 2 *left*). This mirrored our original wizard interface, including the visual and audio feedback to wizards indicating the maximum number of chars that could be entered at one time.

The Discourses interface provided a hierarchical listing of the higher-level discourses (Fig. 2 *right*); rather than typing out what the player says, the wizard selects something that matches what the player *means* (e.g. "Player takes sides by flirting with Grace"). The Discourses interface went though a number of iterations to improve its usability, including incorporating feedback from the docent wizards after the first couple weeks of the show. The current interface represents the discourse hierarchy across three columns that can be navigated through any combination of number keys, arrow keys, and the mouse.

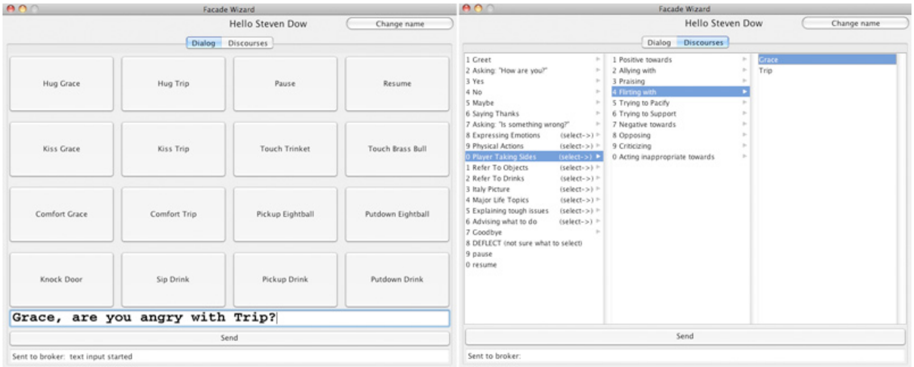


Fig. 2. (left) Dialogue wizard interface provides a text entry field and buttons for indicating player specific gestures. (right) Discourse wizard interface provides a hierarchy of discourses or player intentions, such as "Player Taking Sides->Flirting with->Grace".

3 An Investigation of the Wizard Docents

We initially created the Discourses interface because we believed we could improve on the fundamental time delay between a player speech utterance and the responses by the characters. We wanted to know if wizards could use the Discourse method, and how it compared to the typing interface. Would it be as fast, slower or faster than the other method? How long would it take to learn the discourses? And, how would it affect the experience for players and audiences?

Throughout the eleven-week installation we recorded wizard activity and conducted open-ended interviews with the nine docents, non-technical female art students between the age of 19 and 22. After a short preliminary interview, we trained each wizard to use both versions of the wizard interface and instructed them to create the player experience however they saw fit. During the first 9 weeks of the show—the period where we were not onsite—we allowed the docents to use either wizard interface. During the final 2 weeks, we noticed that most wizards relied on the Dialogue interface, so we forced the docents to learn and use the Discourse interface. We conducted numerous open-ended interviews during the final 2-week period. In this section share some of the insights gained from these interviews, supported by the usage patterns and qualitative sentiments expressed by the docent wizards.

3.1 Wizard Interface Usage

Examining the game episodes for AR Façade for the eleven-week period we can visualize the activity of the wizard docents. There were 106 full episodes—some repeat players—distributed across the nine wizards. The wizards were free to use either the

Dialogue or Discourse interface, and the data shows the former was used more frequently, although W8 used the Discourse interface nearly half the time (Fig. 3). Across all episodes and all wizards, lines of dialog were entered 2.9 times per minute while discourses were entered 0.5 times per minute on average.

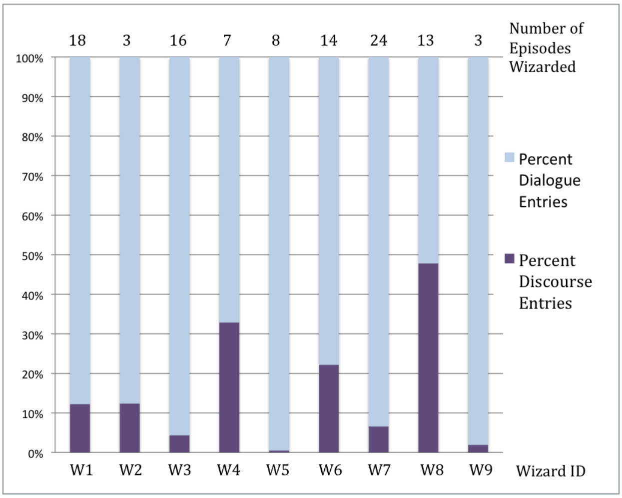


Fig. 3. Wizard activity during eleven-week installation of AR Façade across nine wizard docents

3.2 Wizard Insights on the Two Interfaces

From the wizard interviews, we understood why the wizards generally utilized the Dialogue more than the Discourse interface. There was a consensus among wizards that the Dialogue interface was simply easier; with the Discourse interface "picking stuff out requires more thinking" (W3). According to W4, "it forces your mind to kind of think in a different way, of not just directly translating specifically what they're saying but kind of attributing it to a larger category of emotion or actions. It depends a little more on your interpretation." And W6 explains, "I feel like I have to interpret more what the players are saying. So it's like more involved, like – you have to pay more attention." W8 pointed to one reason to use the Discourse interface: if there was an audio problem, she didn't have to hear the exact phrasing to pick up on the player's intentions.

So the Discourse category selection appears to require more cognitive load, but does it reduce the time delay issue and help the conversation flow better? According to W1, "typing it out takes a little bit longer than searching and clicking," while W4 says "it takes a little longer to sort through them and find the right one then it does to just immediately translate what they're saying into text." Answering this question will require a more detailed conversation analysis, but based on our initial data analysis, we suspect that Discourse selection is faster for more verbose statements and slower for shorter statements.

On the other hand, some wizards reported that the Discourse interface gave them more control over the characters and the course of events. W9 claimed "you could definitely shape the player's experience". W1 even went further, almost describing the characters as puppets: "I'll make Trip talk about the picture again and hopefully he will kind of guide [the player] over." W6 said it was her role "to translate what the people are saying so that Trip and Grace can understand it", anthropomorphizing the virtual characters.

3.3 Wizard Interactions with Players and Audiences

The interviews not only revealed the affordances of each interface, they revealed how the wizards appropriated the system, especially after we enforced the use of the discourse interface. While the Dialogue interface required constant typing and emulation of player speech—especially since the words appear on the player's HMD—the Discourse interface afforded a degree of experimentation. W8 said if there is a "lull in the conversation... I will select something like Therapy, just to offer a little variation. ... because some people would be a little passive in their interactions." For the most part, the wizards were trying to help, but sometimes they got bored and clicked discourses just to spice things up: "I clicked 'have sex' or something because I was hoping that some big explosive thing would happen... I thought that would be fun to see, because Grace seems like kind of an intense chick" (W9).

There were some consequences for this experimentation, as W4 found when she tried using the 'oppose Trip' discourse: "I guess that was too strong of an emotion, because [Trip] kicked him out. haha!" Wizards reported that the system did not always respond as expected, that they wanted more nuance, and that sometimes it felt their actions did not have any affect.

In their role as wizards, the docents learned to become very perceptive of the state of the player. "I can tell when someone feels awkward or when people are getting really annoyed by just like the tone of their voice" (W3). Some wizards were so attuned to how players felt, the emotions were palpable, "I could hear [the player's] breathing pattern and ... I could almost feel her getting uncomfortable and it's kind of bizarre because you can hear the audio and then you can see what they're seeing and I almost started to get uncomfortable!" (W9).

The wizards had different theories on what would be most fun for players and audiences. W7 said she would tell players ahead time "the more you mess with [Trip and Grace] the more entertaining it will be." W1 said she also encourages players to play with it "because a lot of the times people tend to like stay in the middle of the room or they don't get near the objects. ...that way there's a little bit more for me to do in the back too." W3 would flatly reveal her role, even invite players and audiences to see behind the curtain. "I would tell people that I'm gonna be back there.... so that they can get the reaction that they want" (W3).

Some wizards felt it was important to preserve the illusion, as W8 states "I would say the illusion is necessary because it frees up this whole possibility of what could happen if they like slapped Trip or something." According to W6, if players saw the wizards they "seemed kind of disappointed. It's like magic tricks. When you find out it's not fun" (W6).

4 Discussion

While the question of whether the Discourse interface reduced input delay remains to be answered in detail, we learned a great deal by exploring alternative wizard infrastructures. As expected, the Dialogue typing interface did not require as much mental attention, so the wizards initially preferred it. Once we forced them to use the Discourse interface and they starting experimenting and understanding how it worked, the wizards became more attuned to the course of the story and the emotional level of the players and audience members. While the Dialogue interface provided a good design from one standpoint (easier to learn and simpler to use), the Discourse interface engaged the wizards as performers of sorts, enabling them to guide the direction of the experience as they saw fit.

5 Conclusion

In our investigation of the nine non-technical wizard docents that ran AR Façade for 11 weeks, we found that (with training) the wizards became receptive to the Discourse interface where they choose categories of intentionality, not just type in player statements. Although it was more difficult to master, the discourse method gave wizards more control over the course of events. Our wizards reports they would deviate from and actively assert players intentions to create a more appeasing experience for visitors, as well as for their own enjoyment.

References

1. Baum, F.: The wonderful wizard of oz. George M. Hill, Chicago (1900)
2. Beall Center for Art and Technology, `http://beallcenter.uci.edu/`
3. 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 Transactions on Computer-Human Interaction 13(1), 100–133 (2006)
4. Dow, S., Mehta, M., Lausier, A., MacIntyre, B., Mateas, M.: Initial Lessons From AR Façade, an Interactive Augmented Reality Drama. In: ACE 2006: Proceedings of the 2006 ACM SIGCHI International conference on Advances in Computer Entertainment, ACM Press, New York (2006)
5. Koleva, B., Taylor, I., Benford, S., Fraser, M., Greenhalgh, C., Schnadelbach, H., vom Lehn, D., Heath, C., Row-Farr, J., Adams, M.: Orchestrating a Mixed Reality Performance. In: CHI 2001: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 38–45. ACM Press, New York (2001)
6. Mateas, M., Stern, A.: Façade: An experiment in building a fully-realized interactive drama. In: Game Developer's Conference: Game Design Track (2003)
7. Pausch, R., Snoddy, J., Taylor, R., Watson, S., Haseltine, E.: Disney's Aladdin: First Steps Toward Storytelling in Virtual Reality. In: SIGGRAPH 1996. ACM Press, New York (1996)
8. Quasi the Robot, `http://www.etc.cmu.edu/projects/ibi/platform_hardware.htm`
9. Turtle Talk with Crush, `http://en.wikipedia.org/wiki/Turtle_Talk_with_Crush`

Fast Rendering of Large Crowds Using GPU

HunKi Park[1] and JungHyun Han[1,2]

[1] Game Research Center, College of Information and Communication, Korea University
[2] Institute of Information Technology Advancement, Ministry of Knowledge Economy, Korea
{hunki,jhan}@korea.ac.kr
http://media.korea.ac.kr

Abstract. This paper proposes a fast rendering algorithm for real-time animation of large crowds, which is essential for video games with a large number of non-player characters. The proposed approach leaves the minimal work of rendering to CPU, and makes GPU take all the major work, including LOD assignment and view frustum culling, which have been the typical tasks of CPU. By offloading the rendering overhead from CPU, the approach enables the CPU to perform intensive computations for crowd simulation. The experiments show that tens of thousands of characters can be skin-animated in real time.

Keywords: crowd rendering, skinning, instancing, GPU.

1 Introduction

In interactive environments, animation of large crowds is an area of research that has been receiving an increased amount of interest. Especially, it is becoming essential in video games that accommodate a number of non-player characters (NPCs). A good example of NPCs is a herd of land animals.

This paper proposes a rendering algorithm which is implemented mostly within GPU. The performance of the algorithm is excellent enough to render tens of thousands of game characters at real time. The experiment results show that the proposed algorithm is appropriate for applications which render extremely large crowds, and require CPU-intensive computation for the crowds simulation.

The organization of the paper is as follows. Section 2 reviews related work. Section 3 describes the rendering algorithm, which is composed of three phases. Each phase is discussed in a subsection in detail. Section 4 presents the implementation and test results. Finally, Section 5 concludes the paper.

2 Related Work

Many works in the area of crowd rendering have adopted the image-based technique based on impostors [1,2]. Dobbyn et al. [3] proposed a hybrid technique for large crowd rendering, where polygon meshes are used for characters in close proximity and pre-generated impostors are used for distant characters. However, impostors do not support free and fast navigation of the viewpoint.

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 197–202, 2008.

In OpenGL and DirectX, rendering an object requires invoking a draw-call. Suppose an object is repeatedly rendered with a distinct world matrix. Then, repeated draw-calls are needed. In order to reduce the draw-call overhead, various instancing techniques have been developed. The early work in instancing handles only the rigid objects [4,5].

In games, the most popular technique for character animation is skinning [6]. The skinning algorithm is based on a hierarchy of bones, and each vertex in the mesh is assigned a set of influencing bones and a blending weight for each influence. A vertex shader implementation of skinning has been reported in [7], where the bone matrices are recorded in the constant registers. Due to the limited number of constant registers, however, the vertex shader-based skinning is not good for rendering large crowds. Wu presented a pixel shader implementation of skinning [8]. As the animation data are stored in a texture, the algorithm shows good performances in rendering large crowds.

The most recent algorithm for skinning animation of large crowds is based on DirectX10, and makes use of the API, DrawIndexedInstanced [9]. The per-instance parameters are encoded into an array of shader constants, and the array is indexed using the system variable, SV_InstanceID. This is often called a constant buffer technique.

3 Crowd Rendering

The rendering algorithm proposed in this paper leaves the minimal work of rendering to CPU, and makes GPU take all of the remaining work, including level of detail (LOD) assignment and view frustum culling, which have been the typical tasks of CPU. By doing so, CPU can be devoted to expensive computations for crowd simulation. The rendering algorithm consists of 3 phases, as illustrated in Fig. 1.

3.1 1^{st} Phase

For a character, an AABB (Axis-aligned bounding box) is pre-computed such that the AABB is tight but large enough to bound all animations of the character. In the current implementation, we have two characters, zebra and gnu, and therefore use two AABBs.

Let us denote the diagonal length of the larger AABB by l. See Fig. 2. The six sides of the view frustum are expanded by $l/2$. Then, each character is reduced

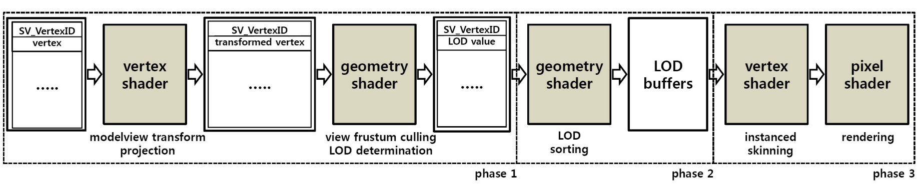


Fig. 1. Flow chart of the rendering algorithm

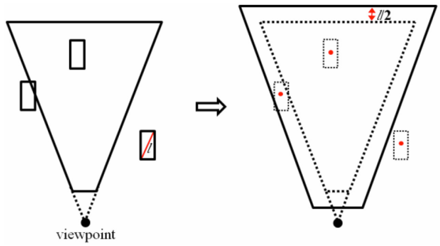


Fig. 2. View frustum culling

into a vertex, the position of which is the center of the AABB. The vertices fill the vertex buffer, and the primitive type is defined to be point list. For each vertex, its ID is automatically generated as a system variable, SV_VertexID. The vertex shader takes a pair of a vertex and its ID, one at a time, as shown in Fig. 1.

The vertex shader transforms each vertex into the clip space, and the view frustum into a $2 \times 2 \times 1$ cuboid. The transformed vertex is culled against the view frustum by the geometry shader, i.e. is tested if it is inside the $2 \times 2 \times 1$ cuboid. If so, the LOD of the vertex is determined using its depth-value.

The current implementation adopts a discrete LOD technique. Fig. 3 shows 3 LODs for zebra. Gnu also has 3 LODs. In fact, zebra and gnu of the same LOD have similar resolutions.

The geometry shaders output is taken as input to the 2^{nd} phase. (No pixel shader is invoked in the 1^{st} phase.) For this purpose, the new feature of DirectX10, stream out, is used. The vertex ID and its LOD value are output, one at a time, by appending them to an output stream object. Fig. 4-(a) illustrates the output of the 1^{st} phase.

3.2 2^{nd} Phase

A number of instances of zebra or gnu will be rendered at the 3^{rd} phase, using an instancing API of DirectX10. Note that instancing requires a single LOD, i.e.

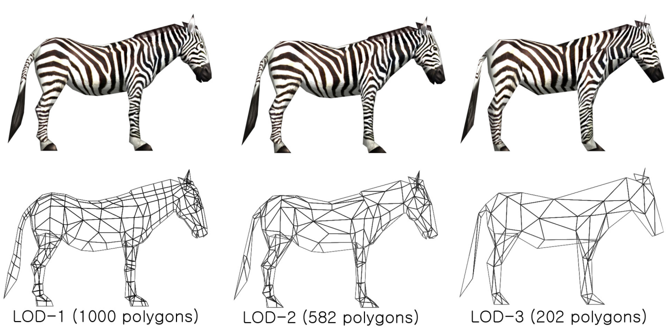


Fig. 3. LOD characters

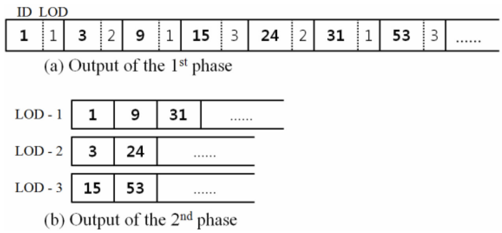


Fig. 4. Output streams

a single-resolution mesh. However, the output of the 1^{st} phase is a mixture of different LODs. The 2^{nd} phase sorts the output of the 1^{st} phase into appropriate LOD buffers such that all of the characters in an LOD buffer have the same resolution mesh. As 3 discrete LODs are used in the current implementation, 3 LOD buffers are created.

An LOD buffer is populated by invoking a draw-call, and we need 3 draw-calls in total. For each draw-call, the vertex shader is passed through, and the geometry shader collects the vertices (reduced characters) that share the common LOD value. Fig. 4-(b) illustrates the output of the 2^{nd} phase.

3.3 3^{rd} Phase

The last phase of the proposed algorithm performs skinning animation for each instance. For this purpose, the constant buffer technique [9] has been adopted. Both of the geometry data and the animation texture remain fixed for the entire animation. In contrast, the instance buffer implemented in the constant buffer is repeatedly updated. After the instance buffer is updated, the 3-phase rendering starts. Let us return to Fig. 4-(b), the output of the 2^{nd} phase. For each LOD buffer, DrawIndexedInstanced is invoked, i.e. we need 3 draw-calls. Each element in the LOD buffer works as an index into the instance buffer. The appropriate LOD mesh is world-transformed, and then skin-animated using the animation frame extracted from the animation texture.

4 Implementation and Test Result

The proposed algorithm has been implemented in C++, DirectX10 and HLSL on a PC with Intel Core2 Duo E6400, 2GB memory, and ATI Radeon HD 2900 XT 512MB. For experiments, two characters, zebra and gnu, are used, each with 44 bones. For both, the finest resolution mesh consists of 1,000 polygons.

Fig. 5 shows the frame rates of the proposed rendering algorithm, where the percentage indicates how many characters are inside the view frustum. For example, a scene of 5,000 characters is rendered at 31 fps when 4,500 characters (90% of 5,000 characters) are inside the view frustum. For analysis, the characters are uniformly spread within the view frustum such that about 1/3 of the

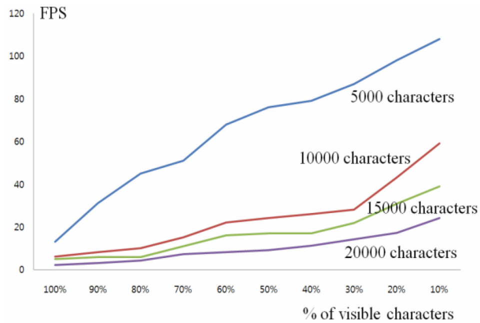


Fig. 5. Performance visualization

Fig. 6. Rendering results

visible characters is sorted to an LOD buffer, i.e. each of the 3 LOD buffers contains about 1,500 characters. Obviously, the frame rates get higher when more characters are discarded through view frustum culling.

Fig. 6 shows snapshots of rendering 20,000 characters. The average percentage of the characters inside the view frustum is 27%, and the average FPS is 15. For crowd simulation, the flocking algorithm of Reynolds [10] is used, but any simulation technique can be integrated with the rendering module.

5 Conclusion

This paper presented a rendering algorithm for real-time animation of large crowds. The rendered scene shows tens of thousands of game characters, each animating in different poses, rendered using just a few draw-calls. In the current implementation, 7 draw-calls are needed in total. The rendering algorithm is implemented mostly within GPU. By offloading the rendering overhead from CPU, it enables the CPU to perform intensive computations for the crowds simulation.

Acknowledgments. This research was supported by MKE, Korea under ITRC IITA-2008-(C1090-0801-0046).

References

1. Aubel, A., Boulic, R., Thalmann, D.: Real-time display of virtual humans: levels of details and impostors. IEEE Transactions on Circuits and Systems for Video Technology 10(2), 207–217 (2000)
2. Tecchia, F., Loscos, C., Chrysanthou, Y.: Image based crowd rendering. IEEE Computer Graphics and Applications 22(2), 36–43 (2002)
3. Dobbyn, S., Hamill, J., OConor, K., OSullivan, C.: Geopostors: A real-time geometry/impostor crowd rendering system. ACM Transactions on Graphics 24(3), 933 (2005)
4. Instancing Sample. DirectX SDK (February 2006)
5. Zelsnack, J.: GLSL Pseudo-Instancing. NVIDIA Technical Report (November 2004)
6. Lewis, J.P., Cordner, M., Fong, N.: Pose space deformations: A unified approach to shape interpoalation and skeleton-driven deformation. In: Proc. SIGGRAPH, pp. 165–172 (2000)
7. Gosselin, D.R., Sander, P.V., Mitchell, J.L.: Drawing a Crowd. ShaderX3 (CHARLES RIVER MEDIA), 505–517 (2005)
8. Wu, O.: Animation with R2VB. ATI SDK (2006)
9. Dudash, B.: Skinned Instancing. NVIDIA Technical Report (February 2007)
10. Reynolds, C.W.: Flocks, herds and schools: A distributed behavioral model. In: Proc. SIGGRAPH, pp. 25–34 (1987)

Physiological Player Sensing: New Interaction Devices for Video Games

Jérôme Dupire[1], Viviane Gal[2], and Alexandre Topol[2]

[1] THIM, Universite Paris 8, 2, rue de la Liberté,
93526 Saint Denis, France
jerome.dupire@univ-paris8.fr
[2] CEDRIC, CNAM, 292, rue St Martin,
75003 Paris, France
{viviane.gal,alexandre.topol}@cnam.fr

Abstract. In traditional video games, players are used to handle well known devices such as a keyboard or a mouse. In a pervasive game, no virtual world exists, since players evolve in a real environment. Hence, new kinds of interaction devices are needed. The physiological approach presented in this paper is one of the most promising solutions. We have build our own sensing tee-shirt on which physiological sensors are sewed. It makes it possible to monitor in real time the players' physiological states with a minimum of mechanical constraints.

Keywords: Wearable Computing, Physiological Sensor, Embedded System, Video Game.

1 Introduction

Our work is part of a national funded project named PLUG (PLay Ubiquitous Games). The aim of PLUG is to propose a new way for people to visit a museum. The base idea of this novel approach is to turn agreeing visitors into players and the museum into a game are(n)a.

Each player will be given a mobile platform, such as a mobile phone or a Personal Digital Assistant (PDA). This device makes it possible for the player to receive information from the game server. Traditional interactions between players and the game will be performed through this device, which will be linked to the game server by a Wi-Fi connection (Fig.1). The other technological gaming elements, such as RFID tags or beacons, will be disposed in the museum. According to the game scenario, players will have to interact with them in order to complete various challenges.

In this project, our work package consists in proposing a new kind of interaction in the game environment. The idea is to sense the physiological state of each player and to feed the game engine with these new signals. Hence, for our game, feelings, together with geolocalization, become one of the most important interface between the player and the game, making the usual gestual (through standard gaming devices) and visual interface (through screens) unusable.

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 203–208, 2008.

Our ultimate goal is to bring up a precise psycho-emotional state of the player. Although this approach is new in video games, we are however guided by studies relating to emotions. Indeed, psychophysiology tries today to reveal human emotions according to several biometric signals. It also studies the effects of the affect on human physiology [1].

In the next sections of this article, we present the work accomplished until now. It relates mainly to the sensors and their assembly on the player. We also detail the test protocol set up as well as the first results obtained. Lastly, we conclude by explaining the next stages that should transform this preliminary work into a central component of the middleware developed for PLUG.

2 Hardware Description

Each player will be equipped with specific sensors. We chose to detect the following physiological variables: the electric activity of the heart (EKG), the muscular electric activity, the skin temperature (ST), the galvanic skin response (GSR), the breathing rate (Br), the blood pressure (BVP) with the method described in [7]. In addition to those sensors, we have associated a tri-axis accelerometer whose purpose is to raise ambiguities when dealing with close physical or physiological states. For example, does the heart rate increase because the player started to run or because he/she is stressed?

The Br sensor is based on a variable resistor whose value depends on its physical deformation. The electrodes used for the EKG are Red Dot (3M). For the final sensor tee-shirt, they will be replaced by textile surfaces integrated into the sleeves, as in [4]. The temperature sensor is a LM385, sewed so that it is in contact with the skin. The accelerometer is a ADXL330 (from Analog Devices Inc.). The GSR sensor is integrated in a glove worn by the player.

The set of sensors is connected to the ATMEGA168 (ATMEL) microcontrolor of an Arduino board [2]. We chose this particular board for our tests for two main reasons. Firstly, because the GUI and the progamming language close to C++ make it very easy to use. Secondly, because a great community of developers exists around this material and many shared examples and projects can be found.

The link between the Arduino board and the distant data processing system is made by a point-to-point ZigBee radio connection implemented through XBee modules (Maxstream). Here also, this particular material was chosen for its simplicity but also for its already tested use in contexts close to ours, like house automation or health monitoring [3].

The communication between the embedded ZigBee module and the PC is made through a FT232 (FTDI) interface which makes it possible to set a virtual serial port.

All sensors and electronics are sewed on a cloth, making it very easy for the player to literally wear the sensors. The tee-shirt textile is stretchable in order to fit various players' shapes and to ensure that the electrodes are well plated on the skin. Figure 1 shows the various zones reserved for the various sensors on the tee-shirt. The choice of these various zones was made by seeking the best compromise between an ideal placement for collecting strong signals and the least possible constraining or awkward placement for the player.

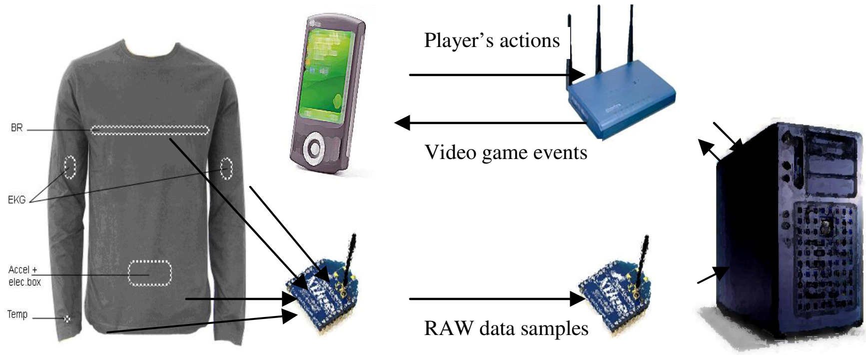


Fig. 1. Position of the various sensors on the tee-shirt, wireless Zigbee and Wifi networks

Moreover, the assembly is made in such a way that wires can slide while arms are moving. A strong constraint of the device is that it must be sufficiently light and discrete to be installed quickly on visitors of the museum. Moreover, it must allow a total liberty of action for the player, without any handicap of weight or handiness. Any exemption from this rule could lead visitors to avoid playing the game or to give up during a game session.

Sensor	Physiological measures	Emotions	Comments
EKG	Increase Drop	Anger Fear Joy Sadness Love Disgust	With some differences
Br	Long expiration Wide range	Anger Joy	
ST	Increase Drop	Love Anger Fear Disgust	
GSR	Increase	Fear	
BVP	Increase Drop	Anger Joy	

3 Experimentation

In the field of psychology, there is a great number of works relating to physiological measurements of psychological phenomena [6], [7], [8]. Our purpose is not to remake these experiments. We want to rely on those works and show that our device answers some of the criteria known and validated by the professionals of this field. To do so, we test users in a passive mode. In a second stage to come, the interactivity will be integrated to achieve our goal of emotionnal driven video games. To do so, we will have to build the emotional identity card of the player with which it will be possible to imagine some new game designs.

The following table gives some examples of feelings according to physiological variations [9]. The combination of various measurements make it possible to distinguish the type of emotion. This distinction is reinforced by the accelerometers.

Our test bench is composed by the sensors carried by the user, but also by a camera and a microphone that record the user and finally by a video submitted to the user. The camera and the microphone make it possible for this first phase of tests to put in correspondence the physical behavior (body, verbal,…) and the physiological behavior of the user. The movie watched by the user is an edition carefully timestamped of still images, sounds, movie extracts that are (or not) emotionnally charged. It thus includes calm and key periods.

The user wears the tee-shirt and a mitten on which sensors take place. Measurements are taken remotely thank to the zigbee module. In order to avoid the emotional parasites, the user is isolated and installed comfortably in a room slightly enlightened in front of a screen. The screen includes an almost invisible camera and a microphone. The film is projected during a fixed duration (around 26 minutes).

The perception and emotional reactions being different from one user to the other, we decided to test two groups of adults: a female group and a male group in the same proportion. A set of 10 people currently lends itself to our experimentation. This phase enables us to adjust the device. The results are stored in files and will be analyzed at the end of the process. A real time visualization tool of the sensors signals confirms that the participants reactions seem in conformity with what was awaited. For example, while degusting images were displayed on the screen, the remote visualization tool depicted heart activity and a skin temperature falloffs (Figure 2.). The complete analysis and filtering of the results will be decisive for the choices of the significant elements to retain.

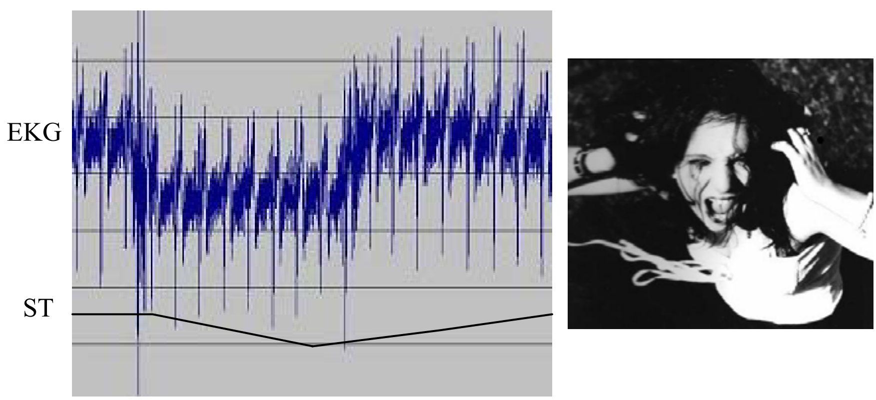


Fig. 2. EKG and ST curves while the right picture is displayed with a scream of terror sound

4 Discussion and Future Work

Taking into account sensors values in a video game can be done in two different ways. The first and most simple one consists in the use of the raw data by the game engine. It will thus modulate in real time the scenario according to values of one or more physiological variables. For example, the aiming cross in a FPS game can shake increasingly when the heart beat or the galvanic skin response are getting to high.

The second manner to deal with signals is to try to interpret the variations of the whole or a part of recorded physiological variables, in order to bring up a more precise psycho-emotional state and send according events to the game server.

This second way of dealing with emotions is much more interesting and is our goal. It would make it possible to provide to the game designer some new tools which could help :

- to conceive and/or validate its traditional game design by finding out if the user really senses the action as it was imagined,
- to develop an adaptive scenario based on the emotions of players,
- to simply take into account the user's emotion like any other traditionnal way of interacting.

First feedbacks from the tests show that all sensors do not work properly during the sessions lengths and should be tuned before a second bunch of test. Some artefacts appear once in a while on curves and may be caused by some interferences between sensors. This problem should also be seriously considered.

Our work will continue with the analysis of signals. Recorded signals should be imported in MathLab in order to ease the determination of the right computations needed to extract feelings. If this step succeeds, we will then code the right filters within our visualization tool to trigger feeling events. The last step before the game designer can truely take into account the player's emotion in its scenario will be to feed the game engine with the feeling events.

Acknowledgments. The authors would like to thank J.F. Lambert (Univ.Paris 8), for his help and welcoming in his psycho-world. The PLUG project is funded by the French national agency for research (ANR).

References

1. Berntson, G., Caccioppo, J., Tassinary, G.: Handbook of psychophysiology. Cambridge University Press, Cambridge (2007)
2. The Arduino project, `http://www.arduino.cc/`
3. Dagtas, S., Pekhteryev, G., Sahinoglu, Z.: Multi-Stage Real Time Health Monitoring via ZigBee in Smart Homes. In: Proceedings of AINAW, May 21 - 23, vol. 2 (2007)
4. Finni, T., Hu, M., Kettunen, P., Vilavuo, T., Cheng, S.: Measurement of EMG activity with textile electrodes embedded into clothing. Journal Physiol. Meas. 28, 1405–1419 (2007)
5. Gross, J.J., Levenson, R.W.: Emotion elicitation using films. Cognition and Emotion 9, 87–108 (1995)

6. Nasoz, F., Lisetti, C.L., Alvarez, K., Finkelstein, N.: Emotion Recognition from Physiological Signals for User Modeling of Affect. In: Brusilovsky, P., Corbett, A.T., de Rosis, F. (eds.) UM 2003. LNCS, vol. 2702, Springer, Heidelberg (2003)
7. Sorvoja, H.: Noninvasive blood pressure pulse detection and blood pressure determination. Oulu University Press (2006)
8. Villon, O., Lisetti, C.: A User Model of Psycho-Physiological Measure of Emotion. In: Conati, C., McCoy, K., Paliouras, G. (eds.) UM 2007. LNCS (LNAI), vol. 4511, pp. 319–323. Springer, Heidelberg (2007)
9. Scheff, T.J.: A Taxonomy of Emotions: How Do We Begin (2005), `http://www.soc.ucsb.edu/faculty/scheff/47.html`

Hitch Haiku: An Interactive Supporting System for Composing Haiku Poem

Naoko Tosa, Hideto Obara, and Michihiko Minoh

Academic Center for Computing and Media Studies, Kyoto University
Yoshida-Nihonmatsu-cho, Sakyo-ku, Kyoto, 606-8501 Japan
{Tosa,Obara,Minoh}@media.kyoto-u.ac.jp

Abstract. Human communication is fostered in environments of regional communities and cultures and in different languages. Cultures are rooted in their unique histories. Communication media have been developed to circulate these cultural characteristics. The theme of our research is "Cultural Computing", which means the translation of cultures using scientific methods representing essential aspects of Japanese culture [1]. We study the reproduction of a traditional Japanese Haiku by computer. Our system can abstract an essence of human emotions and thoughts into a Haiku, a Japanese minimal poem form. A user chooses arbitrary phrases from a chapter of the essay "1000 Books and 1000 Nights" [2]. Using the phrases chosen by the user, our system generates the Haiku which includes the essence of these words.

Keywords: Poem, Haiku, Haiku generation, Art, Interactive art.

1 Introduction

Haiku is a Japanese classical poem style with minimal length of five-seven-five characters 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].

We developed a new interactive system, "Hitch Haiku", which supports a user for composing a Haiku. The user inputs some words into the system, and the system composes phrases consisting of five-seven-five characters which most fit with the user inputs. The system is called Hitch Haiku as it generates a Haiku "hitching" the phrases chosen based on the user inputs. If the user does not like the composed Haiku, the user can modify the Haiku and make the system to learn the composition of better Haiku. Hitch Haiku is one of the automatic poem generation systems. In 1959, Theo Lutz developed a system of poem generation for the first time [5]. The system only showed words at random based on grammatical rule, and could not generate a poem in its real meaning. 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

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 209–216, 2008.

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.

In the field of Interactive Art or Game, the quality of contents is important [7]. But in these cases only simple techniques have been used. On the other hands, in the field of AI, many researchers have been using various kinds of technologies to find some relations among input words/phrases by users and to compose answers in relation to these inputs [8]. 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 as an editor and a philosopher, Matsuoka fund four several basic forms called "*Thoughtforms*," that exist as basic forms of relations among things [9]. By tracing and re-constructing some relations indicated by *"thoughtforms,"* Tosa and Matsuoka created an art work called *"i.plot"*, which displays dynamically hidden relations and contextual emergences of English, Chinese Character, and so on [10]. As this technique can re-construct interesting relations and enable to generate Haiku poems, we apply the technique to our system. Furthermore in our system we included the learning function. If the users do not like the generated Haiku based on the users inputs they can modify the Haiku. The whole process is observed by the system and it learns method of generating better Haiku using these examples. The more the users play with our system, the better relations among words/phrases our system learns, and the better Haiku our system can generate.

2 Process of Generation

Our system generates a Haiku according to the following process (Fig. 1). We describe the detail of the process below.

(1) A user chooses arbitrary phrases from a chapter of a famous Japanese essay called "1000 Books and 1000 Nights", which introduces more than 1000 books covering many genres from all over the world [2].

(2) The system carries out a syntactic analysis for each of the phrase and detects a basic form of noun or a verb from each phrase.

(3) Then the system composes a phrase of a Haiku by adding a special propositional particle called *"Kireji"*, which not only separates a Haiku into three phrases but also gives each phrase a Haiku-like feeling.

(4) In addition to these phrases the system tries to generate new phrases so that the combination of the phrases would expand the imagination of a reader/listener. For this the system uses six types of databases; Haiku thesaurus, Kigo thesaurus, idiom thesaurus, case frame of onomatopoeia, thesaurus, and case frame.

(5) From these databases, the system searches the phrase which is most related to the user inputs.

(6) The system scores all phrases using the following weights: 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 related

words, the system sums the score respectively. The system chooses one of the phrases with the highest score, and by presuming the season of the Haiku from the user inputs choose a Kigo using the Kigo database (Fig. 2).

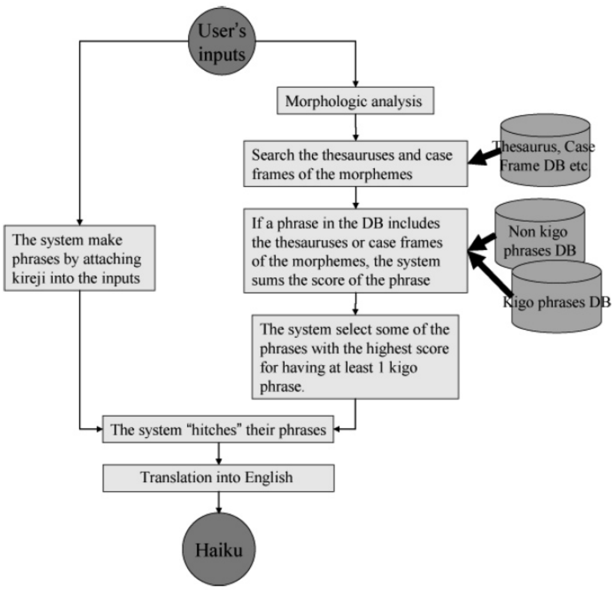


Fig. 1. Haiku generation process

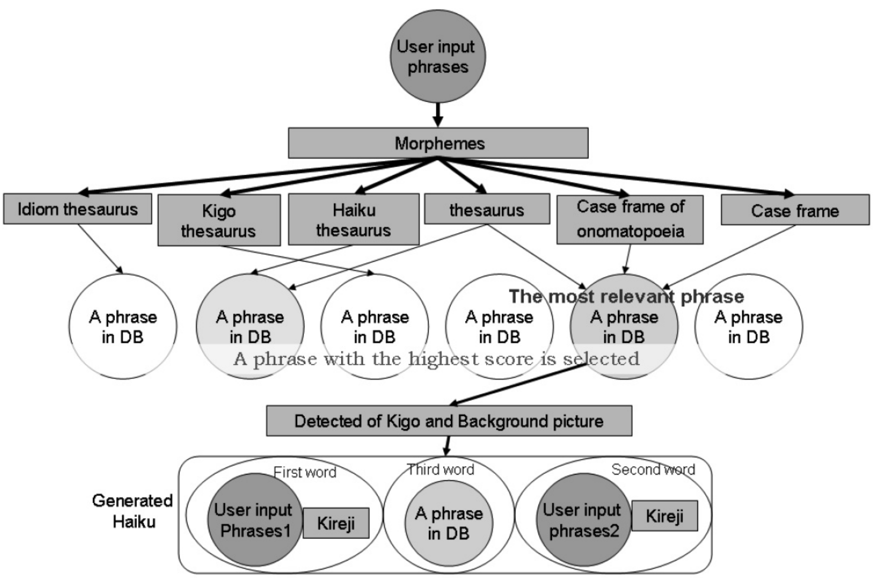


Fig. 2. Selection of the phrase with the highest score

(7) The system translates Japanese Haiku into English Haiku using the translation system, *"Language Grid"* developed at NICT [11]. Language Grid has many dictionaries of translation in many communities. By choosing and using a suitable dictionary of a Haiku community in all dictionaries and by using it, Language Grid translates phrases.

(8) If the user does not like the generated Haiku, the user can modify the Haiku phrases and register these new phrases in the system. We assume that the user inputs have strong relativity with the phrases modified by the use. The system adds the relations between the user inputs and the morphemes of the modified phrases into the database, thus the system learns the new relativity.

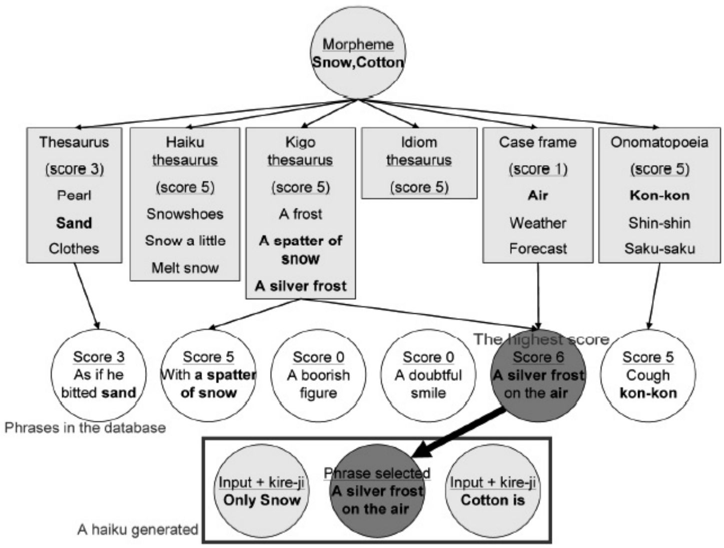


Fig. 3. An example of the phase selection

Figure 3 shows an example of the selection process where two words "*snow*," and "*cotton*" are selected by an user. First to each input word, the most relevant Kire-ji is selected so that two phrases composed from these words would have five or seven characters. As Haiku consists of three phrases, it is necessary to generate one more phrase. For this, the system searches the words related to *"snow"* and *"cotton,"* from the database. Form the thesauruses, *"pearl, sand, cotton"* are found, and from Haiku thesauruses, *"snowshoes, snow a little, melt snow,"* and so on. Then again using the database the system searches Haiku phrases which include these words scores each of these obtained phrases. For example, the phrase, *"As if he bitted sand"* is scored 3 because the word "*sand*" is including in the thesaurus whose score is 3, and *"A silver frost on the air"* is scored 6 because *"a silver frost"* is included in the Kigo thesaurus whose score is 5 and *"air"* is included in the case frame whose score is 1. The system chooses the phrase, *"A silver frost on the air"*, as it has the highest score. Thus the system generates a Haiku by "hitching" the user inputs. If the user does not like the generated phrase "A silver frost on the air," he/she can modify the phrase to new phrase, for example "Like the wind on the air." Then system adds the relations between *"snow"* and *"air," "snow"* and *"wind,"* into the system. (Here, snow and air are the user inputs, wind is a morpheme of the modified.)

Table 1. Examples of the database contents

Database	A Example of Category	Expression
Case frame	fade	sound, taste, part, focus ...
Thesaurus	Expression	Enlightenment, Crying, A facial expression showing familiarity ...
Haiku thesaurus	Love	Relics, Courting, Affection, Loving one another ...
Kigo	Spring	Lunar New Year, First day of spring ...
Kigo	Summer	Summer, Early summer, April, May, First day of summer, Summer-like, Slight heat ...
Kigo	Autumn	Autumn, Early autumn, July of the lunar calendar, August, First day of fall, Lingering summer heat. Autumn-like ...
Kigo	Winter	Winter, November of the lunar calendar, December, Winter solstice, The end of a year, The days of year-end ...
Kigo	New Year	New year, Early spring, The New Year of the lunar calendar, This year, Last year ...
Idiom thesaurus	Love	The whip of love. Autumn breeze blows, Parents often cannot judge the right way to love their children, ...
Idiom thesaurus	Gather	hang out. Many faces, In a crowd, ...
Idiom thesaurus	Walk	turn legs toward, Leave stranded, Extend one ' s trip ...
Case frame of onomatopoeia	Oh	Mouth, Large, Open, Cry ...
Case frame of onomatopoeia	Frank	Taste, Color, Deep, Style, Form, ...
Case frame of onomatopoeia	Hesitant	Attitude, Action Decision, Condition, ...

3 Database

We have prepared six types of databases. The case frame database (with about 31,000 records) describes the relation between verbs and nouns. The thesaurus database (with about 32,000 records) is constructed by classifying words with similar meaning into one category. It is constructed based on a general thesaurus dictionary. The Haiku thesaurus database (with about 2,500 records) is constructed by finding relationship among entries in a general Haiku dictionary. The Kigo thesaurus database (with about 13,000 records) is constructed by setting a Kigo (a seasonal word) as its entity and various kinds of expressions as its contents. The Idiom thesaurus database (with about 1,300 records) contains words as its entries and various kinds of idiom phrases frequently used in Haiku poems. The Case frame of onomatopoeia (with about 8,800 records) contains the relation between an entry word and various kinds of onomatopoeias that are used in relation to the word. The database of user's relation contains the relation between the user input and the morpheme extracted from a phrase modified by the user. The records in this database increase whenever users make any revision to a generated. Some of the examples of the records in this databases are shown in Table 1.

4 Interaction Example

We show an example of Haiku generation by our system.

(1) First if an user logs in the system, he sees a map of the City of Book (Fig.4).
(2) By manipulating a mouse, he can walk around in the city.

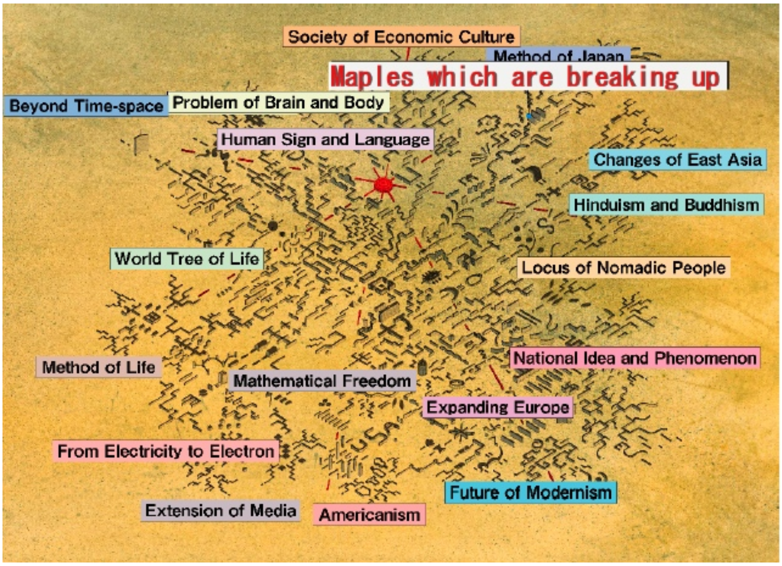


Fig. 4. An example of the map of the City of Book

Fig. 5. Selection of words/phrases

(3) If he finds an interesting keyword of a book at particular points of the city, and clicks the keyword, he sees the title and the author of the book. By clicking again, he sees the essay of the book contained in *"1000 Books and 1000 Nights."*

(4) When the user marks some words on the essay using a pen or an ink brush, the Haiku system generates a Haiku based on the marked words (Fig.5).

(5) If he does not like the generated Haiku, then he can modify and save it.

Figure 6 shows an example of generated Haiku. In Fig. 5, an user selected *"Zen"* and *"ink."* The system chooses the word *"god"* that comes from "*Zen*" using the thesaurus database then chooses the phrase *"The one of the seven gods of Good Luck"* based on the algorithm described above Finally, the system generates the Haiku, *"Even Zen, The one of the seven gods of Good Luck, It's in Ink."*

Fig. 6. An example of a generated Haiku

5 Conclusion

In this paper we proposed an interactive system which supports an user for composing a Haiku. Haiku is a short poem with an abbreviation. Atmosphere and emotion of Haiku is also part of the abbreviation. Based on the algorithm we have developed, we found that sensitive atmosphere and emotion which even we could not expect emerges. We have exhibited this system at ACM SIGGRAPH 2007 and obtained fairly good responses from the visitors [12].

There are several issue for further studies. To increase the contents of each database is one of the important issues. As this system was originally developed for the purpose of Japanese Haiku generation, the process of the translation into English needs a further improvement. At the same time our future target is a system that would support the generation of Haiku based on an interaction between the system and user. So far, what users can do is only to select several phrases/words from texts they like. Although this would be relevant to give a beginner a feeling of Haiku generation, experts would need a system that would really help their creation process. To achieve this target starting from our present system would be a big challenge.

References

1. Tosa, N., Matsuoka, S.: ZENetic Computer: Exploring Japanese Culture, vol. 39(3), pp. 205–211. MIT Press, Leonardo (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. In: Cybernetics, pp. 175–183 (1971)
7. http://digitalmedia.risd.edu/billseaman/workMajor.php
8. 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)
9. Matsuoka, S.: Intelligent Editorial Engineering. Asahi Shinbun Publisher (2001)
10. Tosa, N., Matsuoka, S.: i.plot. ACM SIGGRAPH Emerging technologies. In: Proc. the 32nd International Conference on Computer Graphics and Interactive Techniques, p. 154 (2005)
11. Ishida, T.: Language Grid: An Infrastructure for Intercultural Collaboration. In: IEEE/IPSJ Symposium on Applications and the Internet (SAINT 2006), pp. 96–100 (2006)
12. Tosa, N., et al.: Hitch Haiku. ACM SIGGRAPH 2008 Art Show (2007)

Dome Displays for Educational Games and Activities in the Museum and on the Road

Kerry Handron and Jeffrey Jacobson*

Carnegie Museum of Natural History,
440 Forbes Ave., Pittsburgh, PA 15213, USA
HandronK@CarnegieMNH.Org
PublicVR, 333 Lamartine St., Jamaica Plain,
Massachusetts, USA, 02130
jeff@publicvr.org

Abstract. The immersive Earth Theater at Carnegie Museum of Natural History (CMNH) uses CaveUT and the Unreal Engine for real time virtual tours as part of public and educational programming. Research conducted comparing the theater with desktop displays shows immersion is a powerful tool, helping people create more connections. Content builds on the museum's strengths and includes dinosaurs, an Egyptian temple and a Seneca village. CMNH has tremendous expertise in traditional tours, combining skills of storytelling, communication, leading group dynamics, timing and humor to convey and contextualize the topics. This background is just as critical to great virtual tours as the beautiful, complex and natural graphics.

Keywords: CaveUT, DomeUT, Earth Theater, Educational Games, Informal Education, Dome, Portable, Egypt, Oviraptor, Dinosaur, Virtual Reality, Mixed Reality, Unreal Tournament, Unreal Engine, Immersive, Tour, Constructivist.

1 Introduction

The Earth Theater at Carnegie Museum of Natural History is a partial dome display, 210 degrees wide by 30 degrees high. It seats sixty people within the curve of its screen and is equipped with surround sound. Interactive content for the Earth Theater is based on the CaveUT freeware and the Unreal Engine v2.5 [1]. The software employs a small network of computers driving five projectors which project onto the main screen to provide a single contiguous view of the virtual environment. Prospective effects and spherical warping are accomplished by the VRGL freeware [1], a modified OpenGL graphics library.

While making a presentation in the Earth Theater, the tour guide interacts with the audience while navigating through virtual space using a wireless control device. Just as with a physical exhibit, the tour guide is able to order the presentation and focus on

* Dr. Jacobson conducted this work during his time at the School of Information Sciences at the University of Pittsburgh, 135 N. Bellefield Ave., Pittsburgh, Pennsylvania, USA.

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 217–220, 2008.

individual features according to the needs of the moment—essentially an exercise in storytelling. Unlike a single flat screen, the curve of the Earth Theater provides an immersive space enabling the viewers to be "in the world". It also allows the tour guide to explain more and navigate less while using the full sweep of the view to juxtapose related elements.

Our interactive shows will also display on standard desktop computers, traveling Discovery Dome and a Vision Station™. We will present additional display devices and our most developed applications.

2 The Virtual Egyptian Temple

The Virtual Egyptian Temple [2] has no real-world analog, although it contains elements from existing temples. It correctly embodies the key features of the typical New Kingdom period Egyptian temple in a way that is accessible for an untrained audience. Some differences are made for the sake of presentation. For example, the hieroglyphics are larger than they would be in an actual temple to make them more legible. Nevertheless, the scale and proportions of the spaces are correct; the temple has the minimum required areas (e.g. one Courtyard); the hieroglyphics make the appropriate statements; murals and statuary are in appropriate locations; and so on. The temple is thematically integrated with the museum's Walton Hall of Ancient Egypt collection, with virtual duplicates of some of the artifacts there.

Fig. 1. The Virtual Egyptian Temple as it appears in the Earth Theater

3 OvirapTour

The museum developed *OvirapTour* to accompany the opening of a new exhibit, *Dinosaurs in Their Time*. This three room, virtual mini-museum houses all of the fossils necessary to lead audiences through the evolution of dinosaurs to birds. A fossil and a fleshed-out recreation of *Ovirator philocerotops* begin the story. It continues with the earliest known feathered dinosaurs and earliest known birds. The highlight is a highly detailed model of CMNH's unnamed oviraptorosaur and an ostrich, a very similar modern bird. Part of the challenge for this model was to design it in a

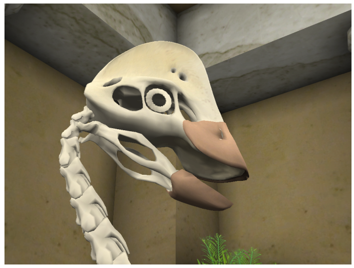


Fig. 2 and 3. Line up of ostrich, ostrich skeleton, *Oviraptor* skeleton and *Oviraptor* in *Ovirap-Tour* playing at CMNH

pipeline for a real time engine despite the complexity required for scientific accuracy. Additionally, it is rigged for animation for use in functional morphology research.

4 Gates of Horus

Gates of Horus is an interactive learning game based on the Virtual Egyptian Temple. The player uses a cordless mouse and a special on-screen(s) cursor to click on individual features of the temple. This prompts the priest in that particular area to explain that feature's meaning. To progress from one area of the temple to the next, the student must answer all of the priest's questions for that area. When the student correctly answers all of these questions, the gateway to the next area of the temple opens, and the student explores and learns in the same way. The student wins the game when s/he

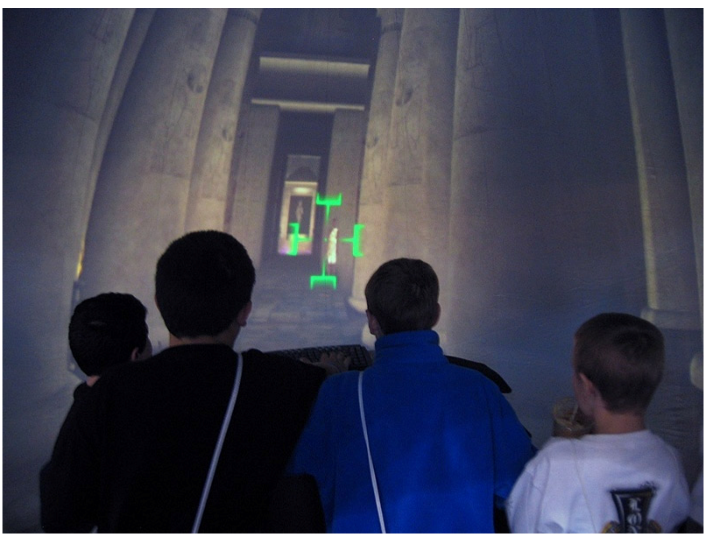


Fig. 4. Children play the learning game, *Gates of Horus*, in the inflatable dome. Image courtesy of Elumenati [3].

answers all of the questions from the priest in the inner sanctuary. Metaphorically, this makes the divine image of the god speak and bring the blessings of heaven to the land of Egypt.

We used *Gates of Horus* for a learning experiment [4], where some students played the game on a desktop computer while others got to (individually) use the entire Earth Theater as their personal display. Those who used the Earth Theater demonstrated more and better knowledge of the temple when asked to *give* a guided tour of the temple. This is one of the very few experiments to actually demonstrate that visual immersion can convey a learning advantage. We believe that the key is to provide a learning activity which truly benefits from the visual immersion and a topic where an inside view of something is informative.

References

1. Jacobson, J., Lewis, L.: Game Engine Virtual Reality with CaveUT. IEEE Computer 38, 79–82 (2005), `http://planetjeff.net/IndexDownloads/Jacobson2005i.pdf`
2. Jacobson, J., Holden, L.: The Virtual Egyptian Temple. In: World Conference on Educational Media, Hypermedia & Telecommunications (ED-MEDIA), Montreal, Canada (2005e), `http://planetjeff.net/IndexDownloads/Jacobson2005e.pdf`
3. Elumenati (2008), `http://elumenati.com`
4. Jacobson, J.: Ancient Architecture in Virtual Reality; Does Immersion Really Aid Learning? Doctoral Thesis, School of Information Science, University of Pittsburgh, Pennsylvania, USA (2008), `http://planetjeff.net/Disseration.pdf`

Game-Based Simulation for the Evaluation of Threat Detection in a Seaport Environment

Allen Christiansen, Damian Johnson, and Lawrence Holder

School of Electrical Engineering and Computer Science
Washington State University, Pullman, WA 99164
allen.christiansen@gmail.com, atagar1@gmail.com, holder@wsu.edu

Abstract. The ability to simulate a seaport environment, including illicit cargo and the sensors designed to detect such cargo, allows the evaluation of alternative detection methods in order to improve security at our nation's seaports. We describe our progress towards this goal. Specifically, we describe our modeling of threats at a particle emission level, modeling of sensors as particle detectors, modeling of the seaport dynamics (e.g, ships, cargo containers, cranes, trucks), and how the particles interact with the various structures and materials in the seaport environment as the cargo moves through the seaport. Ultimately, this simulation will serve as a testbed for the evaluation of sensor network data collection, fusion and decision making for threat detection in a seaport environment.

Keywords: Simulation, seaport, particle propagation, sensor modeling, security, threat detection.

1 Introduction

Advanced approaches to threat detection for homeland security will involve much more non-human intervention via networked sensors and autonomous platforms. Modeling and simulation of these integrated components is necessary for making decisions about deployment and evaluating the intelligence derived from their surveillance. Individual models and simulations of these components exist, but the modeling and simulation of the network of components, considering the physics of the sensors, the physics of the threats to be sensed, and the low-level interaction of the stimulus with the environment, has yet to be addressed.

For example, suppose we suspect a nuclear weapon of mass destruction (WMD) is hidden in cargo container destined for a U.S. seaport. What sensors should be deployed, and where, based on the likely signature of the threat (e.g., gamma emissions), the available sensors (e.g., gamma detectors) and sensor locations (e.g., indoor vs. outdoor), and details of the environment (e.g., intervening metal, stone, wood, soil or water)? Modeling the components of this scenario and simulating their interactions will help determine the proper course of action to maximize the ability to detect and locate the threat. Related approaches to simulating the seaport environment exist [5,7], but they focus on a one-shot computation of detection rates and wait times based on higher-level mathematical relationships among the relevant factors, and they

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 221–224, 2008.

do not realistically visualize the real-time 3D aspects of the environment, which can be instructive in understanding the evaluation of detection methods.

The purpose of this simulation is to accurately model the information flow from the environment to enable the development and evaluation of large-scale, information fusion and data mining techniques to support decision making. Here we describe the modeling and simulation of the threats, sensors and seaport equipment in order to support the goal of evaluation and the automation of the decision-making process.

2 Threat and Sensor Modeling

We are developing a model of threat devices at the particle-emission level using particle tracking within the Torque Game Engine [8] and the Ageia PhysX processor [1] to track particles emitted from a point in all directions (as depicted in Fig. 1). We track particles and model their interactions with various materials in the environment at the particle level (e.g., refractions and absorptions, as well as reflections). Sensors are also modeled at a low-level, and the collision of particles with sensors simulates the level of activation of the sensor. These sensors readings (simulated or real) contribute to the flow of information from the seaport environment.

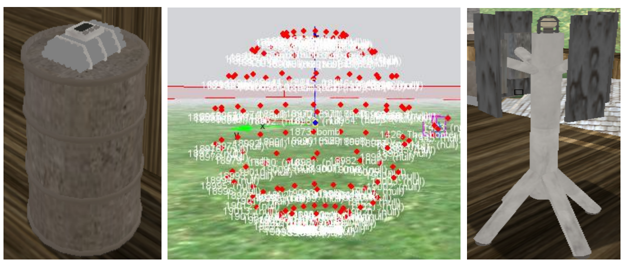


Fig. 1. Threats (left) are simulated as omni-directional particle emitters (middle), and sensors are simulated as particle detectors (right)

The main challenge with modeling these types of threats and sensors is for the particles in the environment to act in a realistic way, despite the computational challenges of tracking vast numbers of particles at relativistic speeds. One approach is to use Monte Carlo simulation tools (e.g., MCNP [3]) which exist to accurately handle radiation particle interactions with given materials. Using data generated by such tools will allow us to better represent realistic behavior.

We are using several tactics to address the computational challenges. First, it is unrealistic to model millions of particles in the environment, so we are using a more modest number of larger objects to represent a group of particles. Rather than trying to keep track of lots of little particles traveling and colliding all over our scene we are instead launching much larger (anywhere from 20 to 100 times the original size) objects. A similar approach has been taken to real-time sound propagation in games [6]. This has the benefit of drastically decreasing the computational load, as well as allowing us to simulate a much larger number of particles active in our environment. Second we are trying several "delivery systems" (using built-in particle emitters, projectiles, third party physics engines, etc.) to determine which will be the most system-resource efficient. One of these methods is the Torque Game Engine's (TGE) [8] built-in particle emitter. This required changes to the engine code to allow custom collisions between the particles and the objects in the environment (e.g., reflection, refraction, and absorption depending on particle energy and object material). However, not having enough control over how the particles were being emitted (one at a time vs. waves like we wanted) caused us to look into a different direction.

The second approach we looked at was just launching the particles with the engine's projectile system. This allowed us to control exactly how many objects we were launching and how they were launched. So we used this to launch our particles in a sphere from the desired launch point and set the velocity as high as we could while still being able to collide with everything we desired. In an attempt to decrease some of the computations required of the main processor, we looked at importing a 3rd party physics engine, Aegia's PhysX engine [1], into the TGE. Eventually we were able to have the PhysX card handle all of the collisions, and just update the TGE of their locations. Since we had a separate processor handing collisions, this allowed us to have many more objects active in the environment at once and use the same functionality as in TGE's projectile system, easily creating and launching particles in any direction. Custom collisions were handled using call backs that allow us to define pairs of objects or groups of objects as collision pairs, which prompts the PhysX engine to make these function calls when objects of pairs collide. This allowed us to perform custom collisions between the objects that we desired, but having these callbacks fired for all of the primary objects in our environment might be too much of an overhead to handle. We are still working to clear up some of these issues.

3 Seaport Modeling

For our simulation we have decided to focus on operations performed by ships, cranes, yard hustlers (trucks), pickers, and trains in their movement of cargo. Cargo and modes of transportation were modeled in the Blender open-source 3D modeling tool [2]. All models were based after photos of seaport machinery (mostly high resolution images from seaports and the Navy). This has allowed us to determine some semblance of scale and proportions between the models. Renderings, models, and sources for the various finished components can be found via the project's web site at http://www.eecs.wsu.edu/sgl (the WSU Simulation and Gaming Lab). Fig. 2 illustrates these models in Blender and as viewed in the Torque Game Engine.

After models are finished their basic operations within the game engine are implemented via Torque Script, a proprietary scripting language that takes advantage of hooks within the engine, such as collision events, timed triggers, and basic physical properties (such as mass and elasticity). Implemented functionality includes methods for lifting, transferring, and carrying cargo for all modes of transportation.

The real time aspect is not strictly necessary for experimental results. However, including it provides a blend of simulation and data visualization that opens additional uses of interest to a wider audience of users. For instance an accurate real time simulation could possibly be useful to seaport workers in determining security holes introduced after sensors are disabled in an accidental collision or during adverse weather conditions.

Another project with similar aims is the L3DGEWorld data visualization tool for network analysis [4]. This project utilizes OpenArena (a derivative of the Quake III Arena game engine) to reflect network activity, and players actions as network commands (such as blocking IP ranges). While the environments differ considerably, both have a similar goal of allowing users to sift through sizable amounts of real-world data via a player. One lesson we take from this project is to abandon realism at times in favor of usability (such as providing a HUD or aerial view upon request).

Fig. 2. Models of the ship, cargo containers and crane in Blender [2] (top three). Seaport models viewed in the Torque Game Engine [8] (bottom two).

References

1. Ageia Physx Processor, `http://www.ageia.com/physx`
2. Blender, `http://www.blender.org`
3. Brown, F., Barnett, N.: A Tutorial on Using MCNP for 1-Group Transport Calculations, Los Alamos National Laboratory (July 2007)
4. Harrop, W., Armitage, G.: Real-Time Collaborative Network Monitoring and Control Using 3D Game Engines for Representation and Interaction. In: VizSEC 2006 Workshop on Visualization for Computer Security, Virginia, USA (October/November 2006)
5. Koch, D.: PortSim: A Port Security Simulation and Visualization Tool. In: Proc. of the 41st Annual IEEE Intern. Conf. on Security Technology (October 2007)
6. Raghuvanshi, N., Lauterbach, C., Chandak, A., Manocha, D., Lin, M.: Real-Time Sound Synthesis and Propagation for Games. Communications of the ACM 50(7) (2007)
7. Sekine, J., Campos-Nannez, E., Harrald, J., Abeldo, H.: A Simulation-based Approach to Trade-off Analysis of Port Security. In: Proc. of the 38th Conf. on Simulation (Winter, 2006)
8. Torque Game Engine, `http://www.garagegames.com`

BCI for Games: A 'State of the Art' Survey

Anton Nijholt

Human Media Interaction, University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands
anijholt@cs.utwente.nl

Abstract. Brain-Computer Interfacing (BCI) has found applications for disabled users. Progress in BCI research allows looking at applications for 'abled' users. For these applications users have other demands, and they will be critical about devices that limit physical movements and that require long periods of training. Prototype BCI applications now appear in the domain of games and entertainment that aim at adapting and controlling a game using brain signals in addition to traditional physical and mental abilities.

1 Introduction

Brain-computer interfacing (BCI) has seen progress in the medical domain, for example for prosthesis control or as biofeedback therapy for the treatment of neurological disorders [1]. Due to this progress we now see attempts to introduce BCI outside the medical domain [2,3,4]. Presently, BCI in human-computer interaction research assumes that devices to measure brain activity patterns can be non-invasive, similar to offering the user a mouse, a keyboard, a camera, a joystick or a Wii-like controller. The most common way to get information from a user's brain is to use an EEG (Electroencephalography) cap. Such a cap is equipped with sensors (electrodes) that detect activity from different brain regions associated with different brain functions (affect, perception, imaging, movement …). Today's non-intrusive EEG-based BCI technology allows patients to communicate with the outside world and it allows patients to control prosthetic devices. Moreover, it allows 'situational disabled' users, that is, users who are able to use hands, feet, gaze, etc., to add an extra control modality to their abilities. Think of a pilot in a fighter plane, a manager in a crisis situation, or a gamer trying to reach a next level in a game.

2 BCI Developments and Applications

In BCI research and technology for games and entertainment we distinguish developments in the following areas: (1) BCI research that allows us to collect information from brain activity that informs us about the cognitive state of the user. Such information can be used to adapt a game or an interface to the user; (2) BCI research that allows us to develop applications where information derived from brain activity allows us to control an application. In this case brain activity can be evoked 'externally' using certain stimuli, or it can be evoked by the gamer in order to address certain

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 225–228, 2008.

challenges offered by the game; (3) Mainstream BCI research that aims at distinguishing and classifying brain activity in different regions of the brain using pattern recognition and machine learning algorithms in order to get knowledge about the cognitive state of the user or to recognize the user's mental control meant to progress in the game; (4) BCI research that aims at developing hardware and software that allows unobtrusive interfacing with BCI applications; hence, no long periods of training and calibration, and no extensive preparations before a user can start.

We can look at these issues in some detail:

Control by affective state: Measuring cognitive activity allows us to detect what the user is experiencing during specific tasks. Such information allows the interface to decrease or increase task load, or, in the context of games, to increase or lower the difficulty level of the game. There are other ways to infer cognitive processes that provide us information about interest, attention, confusion, frustration, irritation, relaxation, satisfaction, fatigue, boredom, etc. Other behavioral and physiological measures are possible, and in game environments we certainly want and have the possibility to get information from eye gaze, facial expressions, body movements and physiological information other than brain activity (hearth rate, blood pressure, skin conductivity). Dynamically tailoring the game to the affective state of the user allows the application to adjust the information flow and to provide effective and pleasant feedback, keeping the user in the flow of the game [5].

Issuing commands by brain signals: What brain activity can be induced by the gamer in order to control the game? In addition, can the game environment be designed in such a way that at certain game decision moments it evokes brain activity to choose among possible alternatives? Various kinds of brain activity that can be used for game control can be distinguished. We can look at brain activity related to imagining movements, activity that occurs when an anticipated event really happens, activity related to a difficult mental task, or activity that is explicitly and externally evoked in order to guide the user in decision making. For example, imaging a movement of your left foot, imagining a movement of your index finger, or imagining a movement of the tip of your tongue, all these imagery movements lead to distinguishable brain activity in the motor cortex of the brain. Similar activity in similar regions of the motor cortex appears when executing or intending to execute such movements. Hence, there is a natural mapping from imagining movements to having these movements executed in a prosthetic device, a robot or as commands in a game environment. Event-related potentials (ERPs) are evoked by external stimuli to which the brain responses automatically. Other brain activity can be evoked by artificial stimuli such as, for example, flickering lights on a screen. Clearly, games can be designed in such a way that such stimuli are a natural in the game.

Developments in BCI technology: In current BCI research we see the use of EEG caps with 32 until 256 electrodes to measure brain activity. We do not know whether so many electrodes are necessary for useful game applications. We see also the use of EEG in combination with EMG (electromyogram) techniques. Rather than being happy with additional sources of information, BCI researchers often consider influence of muscle and eye movements as artefacts that have to be discounted. We can not expect that current EEG caps will become part of a gamer's equipment. In addition to being expensive, it is not necessarily the case that for games we need up to 256 electrodes measuring activity from various parts of the brain. Moreover, setting up a BCI session takes too much time, requiring application of conductive gel, electrode positioning and clean-up after a session. We now see the development of 'dry-cap' technology, allowing a user to use an EEG cap in a similar way as using a mobile phone, an iPod, a Wii, a microphone or teleconferencing equipment. An issue that certainly should be addressed is that in experiments not all users are able to perform at the same level. That is, not all users are able to imagine movements or perform other mental tasks in such a way that a BCI system is able to detect them.

3 BCI for Games and Entertainment

In the previous sections we already made many references to BCI and games. Maybe, in addition, it should be mentioned that BCI game applications are not that different from BCI medical applications or BCI military applications. Gamers, patients and military are handicapped. That is, the circumstances in which they have to perform challenge their abilities to control the environment and these circumstances can ask for control that can not be delivered by available conventional modalities (speech, gaze, keyboard, mouse …). 'Induced disability' or 'situational disability' are words that are used to describe these circumstances. Everybody, handicapped or not, will meet situations where benefits can be obtained from extra communication modalities.

This is particularly true in games, sports and entertainment situations where users have to compete. There are also other reasons that make games, gamers and the game industry interesting. Gamers are early adaptors. They are quite happy to play with technology, to accept that strong efforts have to be made in order to get minimal advantage, and they are used to the fact that games have to be mastered by training, allowing them to go from one level to the next level and to get a higher ranking than their competitors. Moreover, there are enormous numbers of gamers. Having advantage by being the first to introduce a new type of game or a new game element may bring game companies enormous profits. This certainly is an impetus to invest in research and development in brain-computer interfacing.

Nowadays, when we look at BCI games we are asking for theory that allows us to distinguish and employ activity in different regions of the brain (using machine learning algorithms) and that allows us to map each of these activities to commands that are meant to control or adapt a game. These activities can be evoked because:

- the gamer is experiencing the game, the task and the interface, and gets, among others, frustrated, engaged, irritated, bored or stressed;
- there are external stimuli (visual, auditory, ..) consciously generated by the game to force the user to choose among certain possibilities (i.e., make decisions in the game) or that occur in a more natural way because BCI recognizes that a gamer is interested in a particular event that happens during a game;
- the gamer consciously tries to evoke this activity by performing a mental task; e.g., imagining a movement or doing a mental calculation, leading to brain signals that can be recognized and transformed in such a way that the application is controlled by this imaginary movement.

4 Conclusions

There are examples in which researchers 'play' with potential BCI game applications. These attempts are done to develop knowledge about BCI and sometimes there is the explicit aim to contribute to the development of BCI applications and BCI games. Game-like situations have been designed to illustrate research. User-controlled brain activity has been used in games that involve moving a cursor on the screen or guiding the movements of an avatar in a virtual environment by imagining these movements [4]. Relaxation games have been designed [6] and also games that adapt to the affective state of the user [5]. Many variations have been introduced: controlling Google Earth, playing the game of Pong, balancing an avatar on a rope, navigating in Second Life, etc.

There are many challenges unique to BCI applications in HCI. One example is the inevitable presence of artifacts traditionally deemed to be "noise" in traditional BCI explorations. In our applications, we cannot typically control the environment as tightly as in many medical applications nor are we willing to restrict the actions of the user. Hence, we have to devise techniques that either sidestep these issues, or better yet, that leverage the additional information we have available to us. A particular point of interest is how to fuse information coming from more traditional input modalities (e.g. touch, speech, gesture, etc.) with information obtained from brain activity.

Acknowledgements. This work has been supported by funding from the Dutch National SmartMix project Braingain on BCI (Ministry of Economic Affairs) and the GATE project, funded by the Netherlands Organization for Scientific Research (NWO) and the Netherlands ICT Research and Innovation Authority (ICT Regie).

References

1. Coyle, S., Ward, T., Markham, C.: Brain-computer interfaces: A review. Interdisciplinary Science Reviews 28(2), 112–118
2. Nijholt, A., Tan, D., Allison, B., Millán, J., del, R., Graimann, B., Jackson, M.M.: Brain-Computer Interfaces for HCI and Games. In: Proceedings ACM CHI 2008: Art. Science. Balance, Florence, Italy, April 2008, pp. 3925–3928. ACM Publishing, New York (2008)
3. Nijholt, A., Tan, D., Pfurtscheller, G., Brunner, C., Millan, J., del, R., Allison, B., Graimann, B., Popescu, F., Blankertz, B., Müller, K.-R.: Brain-Computer Interfacing for Intelligent Systems. IEEE Intelligent Systems, 76–83 (May/June 2008)
4. Nijholt, A., Erp, J., van Heylen, D.K.J.: BrainGain: BCI for HCI and Games. In: Proceedings AISB Symposium Brain Computer Interfaces and Human Computer Interaction: A Convergence of Ideas, Aberdeen, UK, April 2, 2008, pp. 32–35 (2008)
5. Gilleade, K., Dix, A., Allanson, J.: Affective Videogames and Modes of Affective Gaming: Assist Me, Challenge Me, Emote Me. In: Proc. DIGRA 2005 (2005)
6. Hjelm, S.I., Browall, C.: Brainball – Using brain activity for cool competition. In: Proceedings of NordiCHI 2000 (2000)

In-Depth Observation of Video Gamers

Jorge Alves Lino and Ben Salem

Department of Industrial Design, Eindhoven University of Technology, The Netherlands
jorge@jorgelino.com, mail@bsalem.com

Abstract. In this poster we present the observations results of video gamers playing Halo 3 on the X-Box console. We report on the data we have gathered thanks to a Noldus Observer system. These data relate to gamer behaviour, game actions and basic biofeedback.

Keywords: Game, gamer observation, Behaviour, physiological data.

1 Introduction

Current research demonstrates that the video game industry has been increasing its presence in the field of Entertainment, more and more overcoming the supremacy that before the Film Industry had, this is for example very clear in the US [1]. The audience of video games is becoming more diverse: gaming is no longer an exclusive entertainment activity by male young adults, while elderly and women are becoming more proactive as video game users and players [2]. Video games are becoming a major part of our entertainment habits. However we feel little has been investigated on the way people actually play video games. It is crucial to extend and refine the observation of video gamers, up to a level where we can draw correlations between gamer behaviour, character actions, and physiological parameters. We are also interested in assessing differences if any between genders, socio-cultural and age groups.

2 Method

We propose to conduct video game sessions, where players should go through a stage of the game. During this playing experience, we will observe simultaneously the gamer behaviour, character actions as directed by the gamer and physiological parameters, and compare the different correlations between these factors. We believe that the observation of galvanic skin conductance and heart rate can provide us an overview of the state of arousal of the player, which combined with the behaviour adopted, can provide us valuable results on dexterity and experience of the gamer. During this process, the fact that we are conducting observations with both male and female subjects will allow us to conclude whether there are gender differences or not.

3 Installation

For these experiments, we used the GameLAB facility of the Department of Industrial Design of the Eindhoven University of Technology.

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 229–231, 2008.

The installation set for the gaming observations includes the following equipment:

Hardware: 42-inch LCD Display, Microsoft X-Box, sound system with 4 speakers and 1 sub-woofer, the Biopac, 3 observing cameras.
Software: Noldus Observer, Noldus Acknowledge, Microsoft Halo 3.

Fig. 1. Our GameLAB installation

Fig. 2. The Biopac Module used for biofeedback observations

4 Observations

Subjects between 20 and 30 both male and female are asked by the test moderator to participate in the survey. On the scheduled appointment, they are individually taken to the GameLAB, where a previously prepared A4 sheet with the game instructions is provided, and they are installed in the facilities and connected to the observation system. Subjects carefully read the instructions in order to have an overview on the purpose of the task within the game, and they are asked to play a 2 minute trial to experience the different instructions. Before starting the moderator fixes sensors on the fingers and ear of the test subjects. While this trial period is going on, the test moderator checks if the observation system is well connected to the subject, and working properly.

Once basic understanding of the task is achieved, and the observation system is checked, subjects initiate a campaign, where they perform as real players. While playing, the following parameters are being observed:

- Physical behaviour (expression, manipulation of remote, body language)
- Character behaviour (the kind of strategy is adopted within the game)
- Biofeedback during gaming experience (Heart rate, galvanic skin conductance)

Using the observation software Noldus Observer allowed us to observe simultaneously biofeedback, namely subjects' heart rate and galvanic skin conductance, and the gaming experience itself. Our lab is also equipped with 3 cameras, with different viewpoints of the player. There was an overall camera view of the whole installation with the player in the middle. There was also a camera pointing mainly at the gamer face to record facial expression variations. Finally a third camera was pointing at the player body to identify body language and game controller manipulations and actions. The software allows a complete synchronisation of the different channels of observation. The different observations can therefore be overlapped in order to detect patterns of activities and highlight correlations between the different parameters we were observing.

5 Results

Regarding the game actions, we observed that male subjects have a direct approach to dangerous situations within the game, by going straight forward to the action point of the game. Female players adopt a defensive and self-protective strategy of their character, by approaching the action point of the game with prudence. Players were inexperienced with this game, but didn't have difficulty in finishing the task they were required.

Regarding the physiological observations and the player's behaviour, the process is still ongoing and the results are still partial, depending on further observations in order to sustain conclusive observations.

6 Discussion

So far, we have observed a clear difference between the male and female players, regarding the way they conduct their character actions. Male subjects engage frontally in the conflict within the game, whereas female subjects assume a strategic approach.
Given the fact that the gamers were inexperienced and they showed no substantial difficulties in accomplishing the task, we concluded that this game has a good quality of design of interaction. Given the gender differences in the observations, we believe it would be interesting if such interaction design was adaptive to the gender of the players. This could be achieved, for example, by creating both male and female characters within the game, in which the role the gamer plays in the game would be related with their own gender (and the behaviour particularities that it involves).

References

1. Davis, J.P., Steury, K., Pagulayan, R.: A survey method for assessing perceptions of a game: The consumer playtest in game design. International Journal of Computer Game Research 5(1) (2005)
2. 2008 Essential facts about the computer and video game industry. Entertainment Software Association, Washington, DC, USA (2008)

BioMedia for Entertainment

Ben Salem[1,*], Adrian Cheok[2], and Adria Bassaganyes[3]

[1,3] Department of Industrial Design, Eindhoven University of Technology, The Netherlands
[2] Mixed Reality Lab, National University of Singapore, Singapore
mail@bsalem.info, adriancheok@mixedrealitylab.org,
aketsocjo@hotmail.com,
www.bsalem.info

Abstract. In this paper we report on a novel form of media we call BioMedia. We introduce the concept and we explain its features. We then present two prototypes we have developed using BioMedia in entertainment.

Keywords: BioMedia, Multimedia, Empathy, Entertainment.

1 Introduction

BioMedia is a new media form that uses some property of living beings, in particular plants, as the medium for communication. Rather than use conventional media, say music, to render some information, BioMedia relies on health, shape, pigmentation or bioluminescence of living organism.

We would like to introduce the concept of BioMedia with our implementation of two entertaining systems. We focus on plants, in fact houseplants, as they have interesting features. Houseplants require continuous attention. Regular watering and soil checks are necessary for a houseplant to thrive and be healthy. Houseplants trigger empathy from their owner as s/he feel a duty to care for the plant. Plants can trigger emotions that go beyond what inanimate possessions can trigger. There are anecdotal evidences, when inquiring about ownership of houseplants, of an emotional and empathy link between owners and plants.

Houseplant form part of a house environment atmosphere and have a role in the overall perception one has of a room. We believe an entertainment value can be added if the plant is the embodiment of some information semantically coupled with users of that room.

BioMedia yield new opportunities for developing a new genre of media. The BioMedia form, content, and delivery are different from conventional media. First as living organism, houseplants cannot be controlled as accurately as electronic actuators. A time delay exists between stimuli and changes of conditions and the reaction from the plant. There is also a necessary minimal duration of stimuli for it to have an effect on the plant. It is difficult to assess with exactitude these parameters as they depend heavily on the individual plant size, light exposure, soil quality etc etc. In fact it is as if each plant has it own characteristics.

* Corresponding author.

S.M. Stevens and S. Saldamarco (Eds.) ICEC 2008, LNCS 5309, pp. 232–242, 2008.

1.1 How Does It Work

Houseplants grow, change shape and react to various stimuli and to the characteristics of the soil they are living on. If taken good care of, houseplants will look healthy and beautiful. If neglected houseplants will wilt and eventually die. As such, there is an interesting opportunity in controlling the houseplant milieu, as it will be reflected in the plant state and appearance. In similar fashion, there is some potentially interesting effect in stimulating a houseplant, as it will result in some physical or chemical changes in the plant. The Concept of BioMedia originates from these opportunities. Described simply, BioMedia is about the control of a milieu or the delivery of certain stimuli as they influence some aspect of the plant.

Figure 1 illustrate the basic principle of BioMedia, a computer system via an actuator controls or influence some aspect of the plant milieu such as the moisture level, the orientation or the temperature. An alternative is for the computer system to influence the plant with particular stimuli, for example a light source.

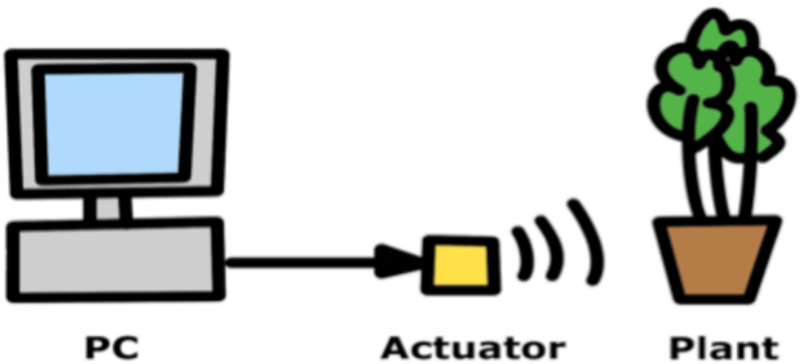


Fig. 1. Principle of BioMedia based on houseplant

In general a BioMedia system will display information over a certain amount of time as plant have adaptation latency. Depending on the plant and what aspects are changed, the adaptation can last anything between hours to months. The plant pigmentation change can take many months (3 months) to occur, as we have experienced with red cabbage (see next section). As a result there is a need to match the speed of change of the houseplant property with the bandwidth and the speed of change of the information displayed with the BioMedia system.

1.2 Relevant Plant Features

To use the plant as a display, we need to provide a control mechanism for one of it features. Several properties of the plant can be used such as health, growth, colour. Other properties can be used for our purpose, gravitropism, thermotropism, and phototropism.

By controlling the amount of water provided to a houseplant, one can control its rigidity and appearance. Wilting would result if a plant is not sufficiently watered. Severe wilting can also be triggered when carefully controlling the watering of the plant. There are some houseplants that are very resilient to water deficit such as cactus and other that are highly adaptive to irregular water availability such as the *Fatsia Japonica*. Plant colour can change depending on the soil conditions. *Red Cabbage* and *Hydrangeas* changes colour according to the soil pH. Other feature of interest gravitropism, is the capability plants have to grow in opposite direction to gravity (i.e. upwards). If a plant is growing on an inclined ground, it will compensate for the slope and still grow vertically (as seen clearly with trees growing on hills). Thermotropism is the capability plants have to react to temperature changes or values. A *Rhododendron* for example curl its leaves when the temperature falls below 2^{o}C. Phototropism is the way plant will grow towards or away from the light source. Thigmotropism is the growth a plant will follow in reaction to a touch or contact. Usually thigmotropism occurs when plants will grow along a surface or around a standing element such as a tree trunk. Resurrection is the last feature we have looked at. It is the capability plants have to dry up and curl in and then blossom and reopen according to water availability (a good example of resurrecting plant is is the *Rose of Jericho*).

1.3 Implementation in BioMedia

Although it is possible to couple any plant feature with a BioMedia system, there are many limitations that need to be taken into account. The main one being the matching between the time duration needed for a plant feature to change and the speed of change of the information being displayed. This matching is necessary to ensure a strong coupling in the user's mind between the information and the plant feature.

We call the time duration for a plant feature change the feature latency, and it can range from seconds to years. At the fastest values range are the rapid plant movement, such as the Venus flytrap whose lobes snap shut when stimulated by a prey. Another

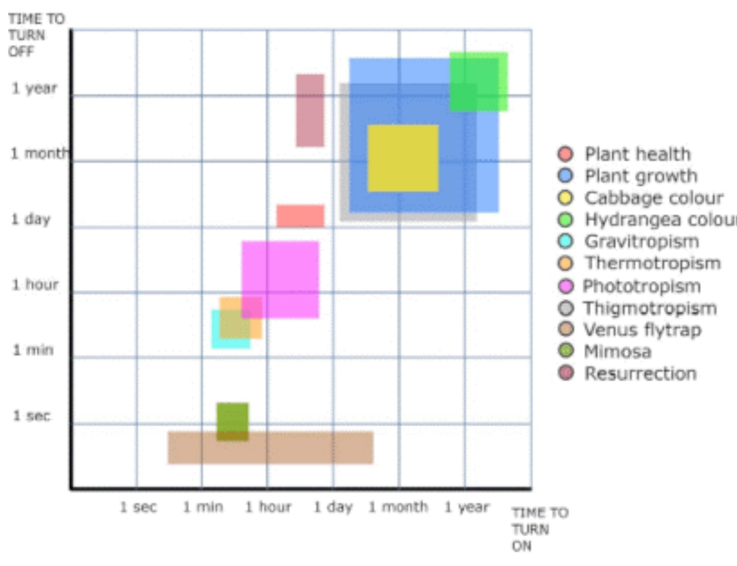


Fig. 2. Range of values for plant feature latencies

feature with short latency is *Mimosa* (*Mimosa Pudica L.*) sensitivity to touch. The mimosa compound leaves fold inward and droop when touched and re-open within minutes. At the other end of the range are features such as plant growth and plant colour changes (such as the cabbage colour). These features have latency that range between months and years (see fig. 2).

Two plant features have attracted our attention, the plant colour and the plant health. We have selected *Red Cabbage* for plant colour changes and we have selected *Fatsia Japonica* for plant health changes.

2 Related Work

There are various projects using the principles of BioMedia. In this section we review only a limited number of these, selecting the closest ones to our proposal. Laughing Lily is a BioMedia system based on an artificial lily that has a carefully controlled blooming (see fig.3). The flower reacts to surrounding sounds. If nobody is talking the flower lets it petals droop [1].

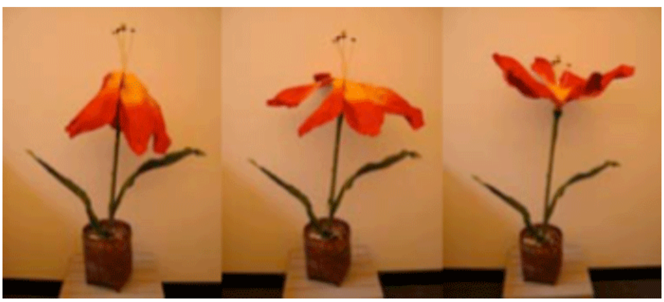


Fig. 3. Phototropism as used in the Infotropism project [1]

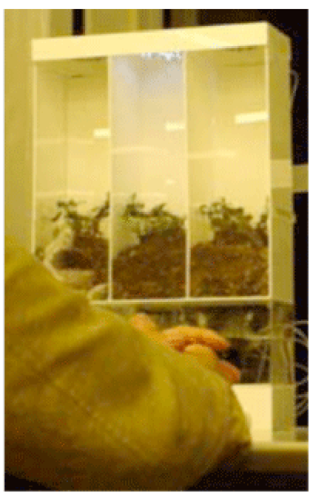


Fig. 4. Phototropism as used in the Infotropism project [2]

Using real houseplants, PlantDisplay is a system where the growth of a plant is controlled by the amount of water given (see fig. 4). In the communication application, the watering of the plant is dependent on the number of telephone calls and emails one exchanges with friends [2].

With the infortropism project, it is the phototropic behaviour of the plants that is exploited to visualise the amount of recyclable rubbish in comparison to regular rubbish. Each time one of two rubbish bins is used, a light is activated. This light activation will gradually induce some plant growth towards it. So if there is more usage of a particular bin, this will result in a growth of plants towards its related light source (see fig. 5). There are also mechanical plants in a remote location that mimic the real plants [3].

Fig. 5. Phototropism as used in the Infotropism project [3]

3 BioMedia Applications

The general direction that we have adopted was one of using BioMedia as part of a lifestyle display. In this scenario, various kind of information could be used. We have made a selection out of information pertaining to relationships, health, bad habits, money, and work. At the implementation level, we have investigated two applications using BioMedia. The first application we have developed with BioMedia is the display of a user's health information. The second application we have investigated is the display of lifestyle information.

In both cases we were interested in using the new media as an entertaining ambient system that connect people with houseplants via an empathy link. In general such system is made up of some sensors connected with the user, either directly to his/her body or indirectly through some personal information, for example a diary. The sensors collect some data related to the user. This data is processed into some information that is fed to the system and relate to the user lifestyle, health or personal habits. The information is then used to control some actuator that has an effect on the houseplant environment. These changes on the houseplant environment are reflected by some changes in the houseplant properties. These changes should not be to closely matching the actuator action; otherwise the control loop is too accurate. We need a loose accuracy as we are designing BioMedia as an entertainment system. As such we want to leave some scope for uncertainty and interpretation. An example that should work is the change of the soil acidity that has an effect on the houseplant pigmentation.

4 Health Display

In the health display we implement a BioMedia system coupling an indication of heart health with the colour of a plant, as outlined in fig. 6.

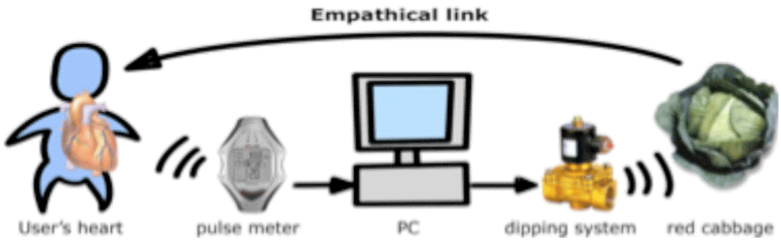


Fig. 6. Health information display system principle

The *Red Cabbage* change colour thanks to a natural pigment called anthocyanin. This pigment has the property of changing colour according to the plant milieu pH from green/blue for acid milieu to red for alkaline milieu. This explains why the plant is know as having different colour depending on the countryside region where it is growing (see fig. 7).

Fig. 7. Red cabbage with either a green (acid) or red (alkaline) natural pigmentation

In this system we have coupled the user's heart health with the cabbage pigmentation. If the user's hearth is healthy the cabbage will be pigmented green and if the user's hearth is not, the cabbage will be pigmented red. This is possible thanks to the *Red Cabbage* pigmentation dependence on soil acidity.

With our proposed system the user would have access to a non-intrusive display of his/her heart health (see fig. 8). The main application would be to warn users of the long-term negative effects of a bad heart health. It should be perceived as a complement to medical care, whereby it would indicate the need to improve heart health, based on a continuous monitoring of some heart parameters.

Fig. 8. Health display system with mature cabbage (note the dipping system)

4.1 Input Data

We have decided to focus on the heart as an indicator of the health condition of the user. Initial idea was to rely on the heart rate (R) of the user. However R is not considered reliable enough as an indicator of heart condition. A better indicator is the heart rate variability (HRV). It is calculated as the fluctuation of time between heart beats. The HRV is the standard deviation of the R to R interval.

An average heart rate R would be around 60 beats per minute (bpm). This implies an HRV of 1.0sec. However the heart is highly adaptive and the HRV of a healthy heart would fluctuate between 0.5sec. to 2.0sec. Typically a low HRV has been associated with a higher risk of heart problems and a high HRV with a healthy heart. Typically a HRV below 7ms would be an indication of a rather unhealthy heart. An HRV above 100ms will be corresponding to a healthy heart (note that the HRV values do not follow a linear progression).

To calculate a user HRV we need to record R for a specific duration over a fixed period of time. In our case we have measured R of a user at night while sleeping. We have taken data samples over a 12 nights period. We discard the first hour of recording to give the user the opportunity to fall asleep; we then collect data over a two hours period. So we end up with data samples related to two hours over a 12 nights period.

Thanks to the control value ranges, we set up a system that waters the *Red Cabbage* with acid water (thus turning the plant green) if the HRV is above 70ms. The system will water the *Red Cabbage* with alkaline water (thus turning the plant red if the HRV is below 70ms.

4.2 System Design

For the purposes of our system we have created acidic water with a pH 3 using Hydrochloric acid (HCl) diluted in pure water. We have also created alkaline water with a pH 10 using Sodium Bicarbonate ($NaHCO_3$) diluter in pure water. Both water as well as pure water (for the control cabbage) were stored in three 20 liter jerrycans. The three jerrycans were located 3mtrs above the cabbages. With such arrangement, water was naturally flowing downwards into tubes. The water flow was adjusted thanks to electrical solenoid water valves. The valves are driven by a control board connected to a host PC (see fig. 9).

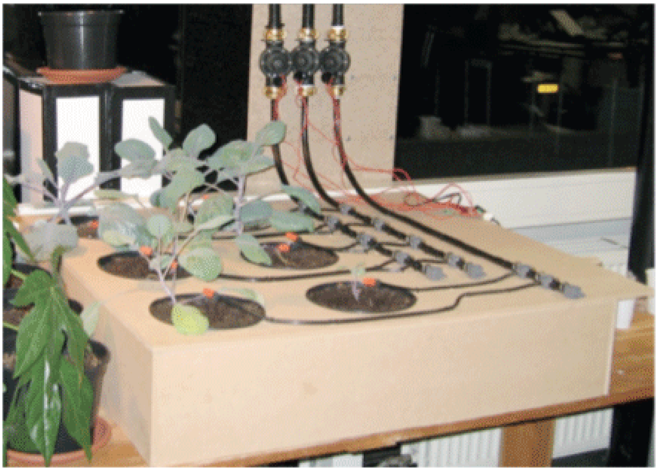


Fig. 9. Health display system (the valves are visible in the back of the installation and the dripping system is visible above each pot)

4.3 Problems with the System

The major problem we have encountered with this system is inherent to the principle of BioMedia. The use of plant as a display device requires some important changes to the plant appearance, in this case a significant change in the plant pigmentation which is a slow process that takes up to three months to start being effective. We have

Fig. 10. Before (top) and after (bottom) watering with acid (left) and alkaline (right) water

known that the feature latency of plant pigmentation is anything between one to several months. However there was no indication in the literature as to how much of a change will be visible. We have encountered the simple problem of the cabbage changing colour but not enough (as illustrated in figure 10). With hindsight we have realised that the acid and alkaline waters must make their way directly into the plant. The dipping display used was water the plant soil. Unfortunately, the plant soil acts as a pH buffer and reacts with the water used. A future version of the system will use a more direct delivery approach, probably based on an intravenous drip system directly connected to the plant main veins.

5 Lifestyle Display

It is a system coupling the smoking habits of staff using a smoking room in our campus with the health of a plant. We couple the usage of the smoking room with the amount of water fed to the plant. Users of the room are explicitly encouraged to discontinue smoking and can see the consequences of their actions via its negative effect on the plant health.

In this system we have used a sensor on the door of a smoking room in our campus to count the number of smokers who are using the room. The plant is watered in inverse relation to how frequently the room is used (see outline in fig. 11). If the room is empty the plant is watered, just enough to bring it back to a healthy state. If the room is being used the plant is left without water so as to wilt.

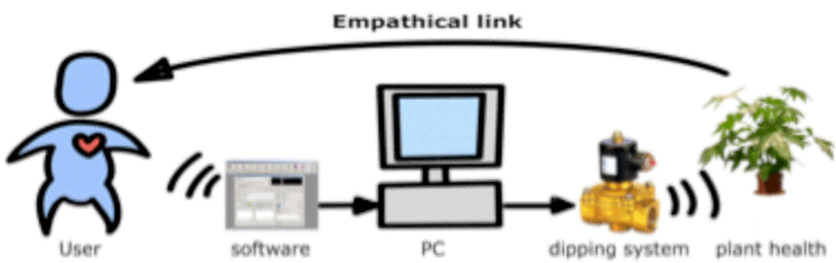


Fig. 11. Lifestyle display system principle

5.1 Input Data

We use a phidget to collect the data from the door sensor. This is fed to a C++ program running on a host PC. Every days, the program predicts the state of the plant by calculating the last 10days watering to assess the soil humidity and hence plant rigidity. Our program then calculates the number of smokers in the room for the last day. This number is used to calculate the desired plant rigidity and thus amount water. The amount of water is then used to control the system valve so as to open and close accordingly.

Fig. 12. *Fatsia Japonica* pronounced wilting without damage

5.2 System Design

Apart from a USB phidget and a host PC our system has a relay that drive an electrical valve. The watering system consists of one big plastic bottle, the electrical valve and a dipping tap. The water bottle is filled with a pH neutral water and positioned at the top of the installation (see fig. 13).

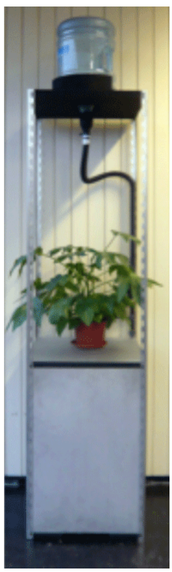


Fig. 13. Lifestyle display as a kiosk in a smoking room

It is important to implement a good control loop, whereby our system waters the plant in a precise enough manner so as to obtain a good effect. To do so we have carefully studied two plants and how they wilt and regain rigidity. We have used the results of this study to evaluate the exact amount of water to give to the plant and what effect this will have (details of this study are beyond the scope of this paper).

5.3 Problems with the System

Surprisingly smokers were not too concerned with the health of the plant. We ran a trial by positioning the system in one of the smoking room in our campus, and we were expecting some reaction from the smokers. Instead, there has been some curiosity followed by indifference. One possible explanation is the denial of one's habit having an effect on the plant; instead believing that it is the others smoking habit that are negatively contributing to the plant health. Further investigations are obviously necessary to find out the truth.

6 Conclusion

Owners of houseplants are normally asked to provide proper light, temperature, water supply and a good soil to ensure a thriving and healthy plant. Thanks to BioMedia, they would also be asked to care for their health or to adopt a balanced and healthy lifestyle. It is obviously a drastic change of purpose for houseplants and an innovative development of media. Although we have only developed two prototypes we hope to continue this work with further systems and to investigate their usage, effect and affect on test subjects.

Thanks to the two early prototypes, we have determined that it is feasible to implement BioMedia systems the way we foresaw it. We also discovered a great potential for adaptation of BioMedia as an entertaining way to continuously monitor health and lifestyle. We hope to further this work and investigate various possible research directions.

Working on BioMedia has been very interesting as we have introduced in our research area things like plants, soil, water and seeds. We have combined electronic components and biological elements: A computer, a micro-controller, and a living plant. There were a lot of issues related to this interfacing. The digital and analogue world of electronic does not mix very well with the biological world of houseplants. One is controlled closely and in real-time. It requires explicit instructions to operate. The other is loosely controlled, with a latency that could last months. It only require various elements (water, soil…) and operates on its own rhythm and pre-programmed genetically behaviour. To make the connection between electronics and biology was an interesting challenge. In our case it is the flow of water that makes the connection. Interesting fact: water the most vital element for life is used to connect an electronic system with a biological organism.

References

1. Antifakos, S., Schiele, B.: LaughingLily: Using Flower as a Real-World Information Display. In: Dey, A.K., Schmidt, A., McCarthy, J.F. (eds.) UbiComp 2003. LNCS, vol. 2864, pp. 161–162. Springer, Heidelberg (2003)
2. Kuribayashi, S., Wakita, A.: PlatDisplay: Turning Houseplants into Ambient Displays. In: Proc. ACE 2006 (poster #40). ACM, New York (2006)
3. Holstius, D., Kembel, J., Hurst, A., Wan, P., Forlizzi, J.: Infotropism: Living and Robotic Plants as Interactive Display. In: Proc. DIS 2004, pp. 215–221 (2004)

Author Index