Experiments on flow and learning in games : creating services to support efficient serious games development

DOI:
10.6100/IR783192

Document status and date:
Published: 01/01/2015

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Experiments on flow and learning in games:
Creating services to support efficient serious games development

Danu Pranantha Dolar

Pranantha Dolar, D.
Experiments on flow and learning in games: creating services to support efficient serious games development
ISBN: 978-90-386-3756-3
NUR: 964
Keywords: serious games / flow / learning / game format / game services / game evaluation / physiological signals / brainwaves / adaptivity / machine learning

Document prepared with LaTeX 2_ε and typeset by pdfTeX (Helvetica and Minion Pro font)
Thesis style design: Martijn Wieling — www.martijnwieling.nl
Cover design: Danu Pranantha Dolar
Printed by: CPI Koninklijke Wöhrmann — http://cpibooks.com/nl/en/

Experiments on flow and learning in games:
Creating services to support efficient serious games development

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr.ir. C.J. van Duijn, voor een commissie aangewezen door het College voor Promoties,in het openbaar te verdedigen op donderdag 15 januari 2015 om 16:00 uur

door

Danu Pranantha Dolar

geboren te Ponorogo, Indonesia

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof. dr. ir. A.C. Brombacher
1e promotor:	prof. dr. A. De Gloria (University of Genoa)
2e promotor:	prof. dr. G.W.M. Rauterberg
copromotor(en):	dr. E.D. Van der Spek
leden:	dr. F. Bellotti (University of Genoa)
	prof. dr. B.A.M. Schouten
	prof. dr. R. Nakatsu (National University of Singapore)
	dr. ir. P.H.M. Spronck (Tilburg University)

The work in this dissertation has been carried out under the auspices of Erasmus Mundus Doctorate Program in Interactive and Cognitive Environments (ICE). This work was conducted towards a joint double PhD degree affiliated with the following partner universities:

UNIVERSITÀ DEGLI STUDI DI GENOVA

TECHNISCHE UNIVERSITEIT EINDHOVEN

Acknowledgements

This PhD Thesis has been developed in the framework of, and according to, the rules of the Erasmus Mundus Joint Doctorate on Interactive and Cognitive Environments EMJD ICE [FPA n°2010-0012] with the cooperation of the following Universities:

Alpen-Adria-Universität Klagenfurt - AAU

Queen Mary, University of London - QMUL

Technische Universiteit Eindhoven - TU/e

Università degli Studi di Genova - UNIGE

Universitat Politècnica Catalunya - UPC

According to ICE regulations, the Italian PhD title has also been awarded by the Università degli Studi di Genova.

"He who has not tasted the bitterness of learning even for a moment, will withstand humiliation from foolishness all his life. Existence of a man is the knowledge and piety." - Imam Asy-Syafi'i q.s.

"The sum total of all the essence of good is to seek knowledge, practice upon it and teaching it to somebody." - Sultaan al-Awliyaa Sheikh Abdul-Qadir Al Jilani q.s.

To my wife and my parents (+in laws)

Acknowledgements

I gratefully acknowledge the support of Erasmus Mundus Joint Doctorate program in Interactive and Cognitive Environments (EMJD ICE) and European Gaming and Learning Alliance (GALA).

This dissertation would not have been possible without the help of my supervisors, colleagues, friends, and family. First of all, I would like to express my appreciation to my promotors: prof. Alessandro DeGloria and prof.dr. Matthias Rauterberg, and my co-promotor: dr. Francesco Bellotti and dr. Erik van der Spek. Throughout the PhD project you have kindly provided me with valuable guidance, knowledge, advice, and support. I also would like to thank the reviewers: prof.dr. Ben Schouten, prof.dr. Ryohei Nakatsu, and dr. Pieter Spronck. I would like to sincerely thank the ICE didactic support: Francesco Grasso and Isadora who had been very helpful during my stay in Italy, and also Miriam for the assistance in the last year of my study. I would like to thank Ellen Konijnenberg and Gaby Jansen for helping me out with all administrative issues during my stay in the Netherlands. All colleagues from EMJD ICE program, Elios lab, and the Designed Intelligence Group, thank you! Luo Cai, Keerati Jittrawong, Andre Possani, Huang-ming Chang, Pongpanote Gongsook, Waqar Baig, Tewodros Atanaw, Ehsan Baha, Boris Takac, Ahn-Tuan Nguyen, Chitra Ayuningtyas, Misha Croes, Idowu Ayoola, Yu Zhang, Kadian Davis, Shadi Kheirandish, Xinwei Wang, Fulya Yalvaç, Maira Brandao, and Veranika Lim, thank you for your help. Furthermore, I would like to thank all my colleagues in Institut Teknologi Sepuluh Nopember (ITS), especially bu Mahendrawathi Erawan for your excellent recommendation, Cak Son, Raras, and Anisah for visiting me during my study. My friends: Dai Ning, Qonita, Reyhan, Reynold, Arya, Xunxun, thank you; PPI Italia, PPI Eindhoven, and KMNU/PCINU Belanda: thank you for having me. I realize that the list of people mentioned in here is far from complete. Nevertheless, I thank everyone that supports me until the end of my PhD.

The last paragraph I keep for people who are simply irreplaceable in my life. My mother, my mother, and my mother, Titik Widayati, thank you for your love, patience and endless prayers; my father, Dolar Yuwono, thank you for giving me plenty of ideas and suggestions. I thank my brother and sister, Arditya Maulana and Fransisca Tiara Saraswaty, for giving me cute niece and nephews. I thank my father (in law), Wen Baochuan for your caring and support, and my mother (in law), Zheng Jingbo for your love and support. Last,

but not the least, I would present my greatest gratitude to my dearest wife, Wen Jing, for always supporting me and taking care of me. You are always by my side and encouraging me during my high and low, thank you! This thesis is dedicated to you!

Danu Pranantha Dolar
Eindhoven, November 1, 2014

Foreword

The work in this dissertation was performed in two European universities: Department of Naval, Electrical, Electronic and Telecommunications Engineering of University of Genoa (UNIGE) in Italy, and Department of Industrial Design of Eindhoven University of Technology (TU/e) in the Netherlands, under Erasmus Mundus Joint Doctorate (EMJD) program. Both universities have different areas of expertise, in which aligning my research between both universities was mandatory. Firstly, in UNIGE, I was involved in European Games and Learning Alliance (GALA) project, in particular to investigate efficient techniques in serious games development under Service Oriented Architecture (SOA) platform. Two key aspects (i.e, the format and delivery strategy) were investigated. To this end, I was involved in constructing a game format, architecture, several services (game features), and modules for game development. Furthermore, I was given a chance to work on physiological signals for adaptivity in games, a huge opportunity to be missed.

On the other hand, during my stay in TU/e, I realized that my work has not yet included the evaluation of services in terms of fun and learning. This is not only important for evaluating the usefulness of services (game features), but also for designing game/game features. Thus, using the services and the system that I have developed in UNIGE, I created two versions of Physics game: with a tutor and without a tutor, and evaluated the effect of the game feature in terms of flow and learning. For the evaluation tools, I adapted eGameFlow questionnaire and created a test set to assess flow and learning, respectively. This work was also intended to clarify the relationship between flow and learning in games.

I hope my contributions in the form of this dissertation can benefit the scientific community, in particular researchers in the area of games and learning technologies. I also wish readers enjoy reading this dissertation.

Contents

List of Tables

List of Figures

Part I

Introduction

Chapter 1

Overview

"We do not stop playing because we grow old, we grow old because we stop playing!" - Benjamin Franklin

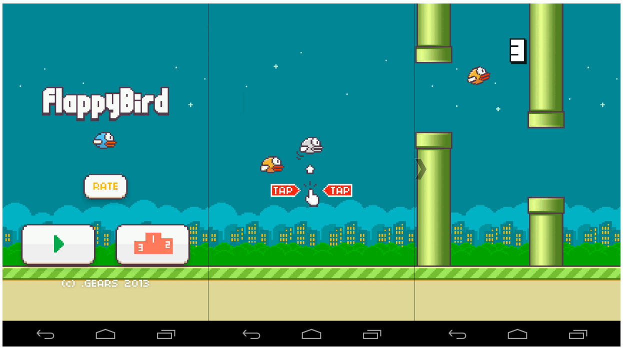
1.1 Introduction

Video games have become a popular form of entertainment and a part of modern culture. Video games gained popularity starting from the golden age of arcade games, coin-op entertainment machines, in the early 1980s with several hit games such as, Pong, Space Invaders, and Pac-Man. Subsequently, various game platforms have been invented, including home game consoles and computer based games. Moreover, with the introduction of the Internet and the web, games have found another platform in the form of online games and social games. The tablets and smart-phones also have gained momentum after the year 2000, and since then we can easily find many mobile game applications for those devices. One of success factors of video games is that they transformed traditional passive entertainment into interactive entertainment, which enables audiences to be actively involved in and influence the outcomes of the games.

One of the most celebrated video games was, without a doubt, Mario Bros (Figure 1.1(a)). Mario Bros was introduced in 1985 by Nintendo and it was well designed in every aspect, from its memorable characters, its intense game levels, and its beautifully crafted story. Mario Bros uses a side-scrolling format and features the journey of two plumbers, Mario and Luigi, in investigating strange creatures that appear in the sewers of New York. The objective of the game is to defeat all of the enemies in each phase. Mario Bros has generated hundreds of spin-offs and inspired games of similar genre.

Flappy Bird, a popular short lived mobile game released in 2013, was also inspired by Mario Bros, although it significantly stripped off the design complexity of Mario Bros, such as game levels and game story. Flappy Bird was a side-scrolling format game where the player controlled a bird, attempting to fly between rows of green pipes without coming into contact with them, by tapping on the screen (Figure 1.1(b)). The player was scored on the number of pipes the bird successfully passes through, with medals awarded for achieving certain scores. Unlike Mario Bros, there is no evolution of play throughout the game as the pipes always have the same gap between them and there is no end to the running track. Although the game is easy to learn, it is hard to master, which probably makes it more interesting. Despite having different levels of design complexity, Mario Bros and Flappy Bird successfully presented challenges to the players. One of the indicators was their popularity and players' remarks on their highly addictive property. One of characteristics of good games is, thus, the ability to capture player's attention for long period and maximize pleasurable feelings or curiosity from the challenges in the games. This involves two of important aspects in games: flow and learning.

(a) Mario Bros video game



(b) Flappy bird mobile game

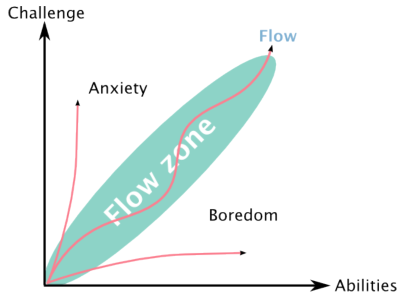
Figure 1.1: Vintage Mario Bros and a retro style Flappy Bird

1.2 Flow in games

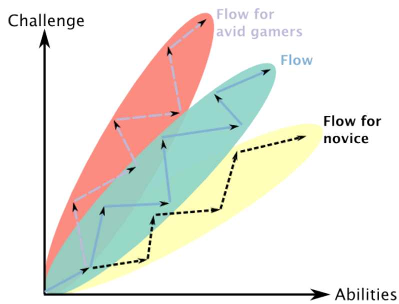
Playing games could engender completely losing track of time and unawareness of the surrounding in a player, feeling of completely engrossed in games that nothing else matters beside overcoming challenges in the games. Some argue that this feeling to be the optimal experience that game designers want to deliver to the players. In the mid-1970s, Mihaly Csikszentmihalyi introduced the concept of Flow, which has since become fundamental to the field of positive psychology (Csikszentmihalyi, 1992). Flow is also called the optimal experience, or being in "the Zone", and it represents the feeling of complete focus in an activity with a high level of enjoyment and fulfillment. This is one of the reasons why people play video games. Other possible reasons may include fantasy, fun, social factor, learning, stories, reward, achievement, and boredom.

Csikszentmihalyi identified eight components of flow: a challenging activity that requires skills, a combination of actions and awareness, clear goals, direct and immediate feedback, concentration at the task at hand, a sense of control, a loss of self-consciousness, and an altered sense of time. Most of today's video games deliberately include and leverage the eight components of flow (Chen, 2007). They deliver instantaneous sensory feedback, offer clear goals and specific gameplay skills to master. However, not all of the components are needed to deliver flow (Csikszentmihalyi, 1992). This means the appropriate mix and match of the components may deliver flow to players. Mario Bros is an excellent example of proper mixture of all components. It provides progressive challenges using game levels, different types of enemies, and multistage game worlds; clear goals i.e. to defeat all the enemies; direct and immediate feedback reflected in the playable character and game world; and easy to master game mechanics. In contrast, Flappy Bird presented constant challenges that are easy to learn but difficult to master and yet it was highly addictive partly because it allowed social competition among players to achieve the highest score. Flow becomes one of the major factors of a good gaming experience, which can be evaluated and compared by measuring flow duration experienced by the players, or by assessing the components of flow and the aspect of social interaction in gaming (Sweetser and Wyeth, 2005).

To maintain the player "in the Zone" as long as possible, the theory of flow emphasizes the balance between the level of challenges and the player's ability to overcome the challenges. If the challenge in the game is beyond the player's ability, the game becomes so overwhelming that it generates anxiety; on the other hand, if the challenge is too easy, the player quickly loses interest (Figure 1.2(a). We have, fortunately, a zone of tolerance for either temporary lack or excess of stimulation (a fuzzy safe zone) where the activity is not too overwhelming nor boring (Csikszentmihalyi, 1992). This provides a range of flexibility in designing appropriate challenges for the player to create an engaging learning curve, which is one of central themes in game design, i.e. content delivery strategy. However, designing a balanced content delivery becomes highly complex as the size of the potential audience grows since different players have different skills and they expect different challenges. Avid gamers or risk takers would like to have extremely high level of challenges to achieve flow whereas novice or casual gamers enjoy a slow pace increase of challenges (Figure 1.2(b)).



(a) flow, boredom, and anxiety



(b) different flow for different types of players

Figure 1.2: Flow

To design a good content delivery for broader audiences, game designer should offer a variety of choices to fit with players' personal Flow Zones. However, populating choices is not only costly but also infeasible. Player and task modeling are frequently used to represent the player's ability and the difficulty of challenges in the form of mathematical models; and subsequently using the models to compute the appropriate level of challenges for the player (Yannakakis and Hallam, 2007; Bellotti et al., 2009a), game AI (artificial intelligence) (Spronck et al., 2004; Yannakakis and Hallam, 2007; Tan et al., 2011). The player's personality could be also included in the model by quantifying personality differences between players, e.g. extraversion trait (Van Lankveld, 2013). Although those approaches are beneficial for optimizing content delivery strategy, the complexity of developing the strategy still remains. This is one of the problems investigated in this dissertation.

1.3 Learning in games

Beside flow, playing games is also about recognizing patterns in the games (Koster, 2013). Once players notice a pattern, they will trace it willfully to see it reoccurs and practice with it until they become fluent and efficient. At a certain point, if players are no longer able to become more efficient, then the games become boring. In contrast, if players meet noise and fail to see any pattern, they become frustrated and give up the game. The theory of the zone of proximal development (ZPD) divides skill development into three zones: *a*) zone where learners can learn independently without any guidance, *b*) zone where learners may learn only under guidance or collaboration with more capable peers, i.e. the ZPD area, and *c*) zone beyond the learners' reach, either without or with guidance (Vygotsky, 1980). In this sense, games may provide guidances (the ZPD area) where they facilitate players in chunking information to be easily digested (in the form of emerging patterns) and provide feedback to enable refinement of knowledge and skills.

Therefore, learning in games for players can be either exercising skills to become more efficient and/or mastering problems mentally (Koster, 2013). First, learning to becoming more efficient involves unconscious skill refinement, quick reflex and judgment. For instance, refining driving skills and strategies from home to office until it becomes an autopilot mode. Second, learning to master problem mentally involves conscious/lateral thinking which is slower compared to unconscious mode. At this stage, learners need to recognize patterns, e.g. other drivers' behaviors. In Mario Bros, the first would be the players need to optimize their strategy in traveling through the game world and defeating the enemies, whereas the second would be learning different types of enemies and various items in the game world to enable him to be successful in the game.

Hence, playing a game is about understanding and solving problems, applying intelligence and wit to overcome the challenges, and practicing skills to become fluent (unconscious processing). It is a cycle of experiencing, observing, abstracting, and experimenting (making decisions) as described by the Experiential Learning Cycle (ELC) shown in Figure 1.3 (Kolb et al., 1984). The continuous learning process in games and their ability to capture player's attention for a long period show the potential of games to assist traditional learning. One of the reasons is that traditional learning often lacks motivation which is an essential ingredient for effective learning (Prensky, 2002); conversely, games are engaging which give strong motives to people to play them voluntarily.

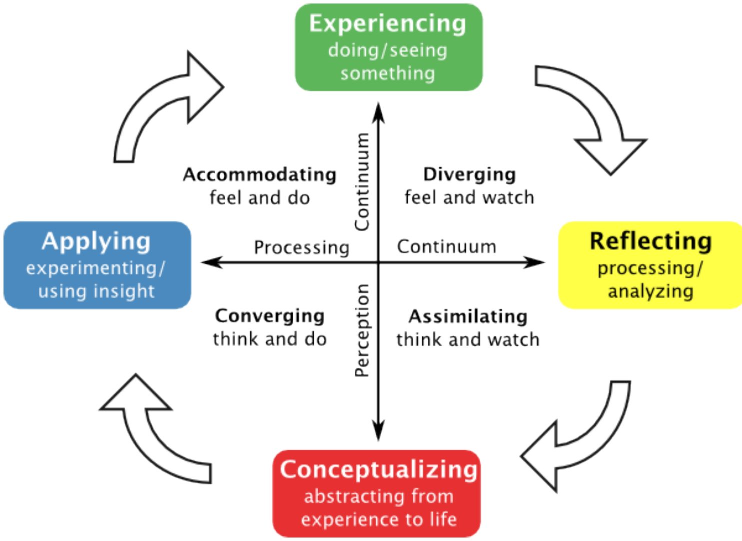


Figure 1.3: Experiential learning cycle (adapted from Kolb et al. (1984))

Thus, researchers in education began to realise that game's innate characteristics, from the intellectual challenge to affording multiple learning styles, may have an immediate role in learning. This led to the introduction of serious games -games that infuse instructions and learning material into play-, becoming more popular (Zyda, 2005). Games, either entertainment games or educational games, can be used to teach conceptual and procedural knowledge, train cognitive skills such as problem solving, enhance motor skills, alter a person's behavior/view, and teach people to communicate and work together (Prensky, 2002; Blunt, 2007; Ratan and Ritterfeld, 2009; Sitzmann, 2011; Van der Spek et al., 2011). Wouters et al. (2009) also mentioned that serious games are a viable means of learning, although there are still plenty of room for improvement.

A literature review revealed that more papers discussed the positive learning outcomes of entertainment games than games for learning (Connolly et al., 2012). This might be due to either the ineffective design of serious games, the difficulties in matching the affordances of entertainment games to specific curricular outcomes (Kirriemuir and McFarlane, 2004), or the limited number of serious games available on the market. As for educational games, they were being used in several curricular areas, commonly in health, business and social issues. However, most studies in educational games were limited to finding whether games were enjoyable and motivating, and lacked in examining the motivational features of serious games in detail (Connolly et al., 2012). Several studies have identified a variety of game features that useful for both engagement and learning in games. Green and Bavelier (2006) found that students' skills can be improved using entertainment games such as attentional and visual perceptual skills. Van der Spek (2011) conducted experiments to provide design guidelines in developing games based on human cognitive processing. Those studies are important to address different features in games that contribute to better learning, such as feedback (Cameron and Dwyer, 2005), varying task difficulty level (Orvis et al., 2007), simulations (Yaman et al., 2008), and the supports for the working memory limitation (Van der Spek, 2011).

Although serious games have gained interests among researchers and practitioners in education, the number of serious games are still limited due to the complexity of the game development. Thus, Serious Games Society is in need of tools to improve the efficiency of serious games development. In this case, beside the delivery strategy explained above, the game format is the key for supporting efficient serious games development (Bellotti et al., 2010). This dissertation was aimed at addressing both game format and delivery strategy as services for improving the efficiency of game development.

1.4 Between flow and learning

Serious games have two goals in nature: learning and enjoyment. Both enjoyment and learning should come hand-in-hand; that is, the better the player enjoys a game, the more motivated the player is to play, the longer the player is being exposed to the learning material, and the better the player learn. In this sense, flow becomes essential to improve learning and to promote exploratory behavior (Kiili, 2005b).

In contrast, Graesser et al. (2009) argued that learning in games should be less entertaining and more about learning. They emphasized the difficulty in learning to allow the player to achieve deep learning. However, later on they found evidence that engagement and deep learning can go hand in hand (D'mello and Graesser, 2013). This is one of problems in serious games, i.e., maintaining the learning process to be engaging and fun while improving the learning outcomes. This necessitates a research into flow and learning in games, in particular to evaluate the services (game features) that may contribute to either flow, learning, or both. This would lead us to the relationship between flow and learning outcomes in games which is now still unclear. Consequently, beside developing the services (format and delivery strategy), this dissertation also aimed at evaluating the effect of a service on flow and learning, and investigating the relationship between flow and learning.

1.5 Research questions

In the field of serious games (SG), there is a clear need for supporting pedagogical authors with methodologies and tools that can support them in providing effective learning (Bellotti et al., 2009b, 2010). Exploring this challenge, there was a number of successful SG (Zyda et al., 2003; Kelly et al., 2007; Sliney and

Murphy, 2008; Mayo, 2007) which used a class - the Sand Box Serious Games (SBSG),with a counterpart also in successful pure entertainment games such as Grand Theft Auto[1] and Oblivion[2] - that tends to provide players with suited knowledge structures for investigating a specific educational domain. SBSG lend themselves well to be defined through an abstract model for facilitating authors in creating adaptive contents (Bellotti et al., 2009b). The abstract model is the environment where knowledge is implemented also through tasks. Tasks embody units of knowledge that have to be solved by the player to progress in the game.

Simple tasks can be realized as instances of configurable software templates that can be easily created by pedagogical authors by simply inserting domain-specific contents, without any need for programming knowledge. This allows creating a wide basis of tasks - also exploiting the User Generated Contents trend that is now popular in TV and multimedia. The subsequent point concerns how to deliver these tasks to the user in a game. In general, two aspects are fundamental when designing tasks: the content and their delivery strategy (i.e. when, where and how they become available to the player or are directly assigned to the player).

Based on this, Serious Games Society (SGS) is initiating the creation of services to support efficient serious games development under a Service Oriented Architecture (SOA). The goal is to provide serious game developers with a repository of a well documented and ready-to-use services (either SOAP or RESTful) usable to develop serious games following the SOA paradigm (Society, 2014). This will prevent researchers and developers from reinventing the wheel since certain functionalities for their game may have already existed as services. This will also support the educators in easy game creation by reusing ready-to-use services for their games. Our work was involved in creating those services which the following questions: *Can we provide open services and modules to support serious games development? What services need to be implemented and how to implement the services? How to evaluate the services?*.

In terms of evaluation, many studies also focused on the effectiveness of game-based learning, in particular knowledge acquisition. For instance, *Supercharged!*, a 3D space navigation game for learning electromagnetism, outperformed guided inquiry group in a knowledge test (Squire et al., 2004). *Re-Mission*, a cancer fighting game in third-person shooter view, improved the players' knowledge on cancer (Beale et al., 2007). Computer science games improved the high school students' knowledge on the concepts of memory in computer (Papastergiou, 2009). A retro-looking computer game on normal distribution improved students' confidence in mastering the basic properties of the normal distribution (Nte and Stephens, 2008). In contrast, a web-based game to teach pediatric content failed to surmount computerized flash card

[1] http://www.rockstargames.com/sanandreas/
[2] http://www.elderscrolls.com

(Sward et al., 2008). A competitive game for learning the functions of human heart failed to improve students' performance, unless feedback was included (Cameron and Dwyer, 2005). However, only few experiments were conducted in controlled settings which weaken the evidence. In addition, some experiments overlooked learning measures, such as knowledge tests. Therefore, several literature reviews emphasize the need of rigorous controlled experiments to provide more evidences on their effectiveness (Connolly et al., 2012; Vogel et al., 2006).

Moreover, several qualitative models have been proposed to both improve and evaluate the learning process in serious games, such as the Game Object Model (Amory, 2007), the Experiential Gaming Model based on ELC and the flow framework (Kiili, 2005a), and the Scaffolding Model (Obikwelu et al., 2013). However, there is little study of the use of those models to evaluate features in games that may contribute to flow and learning. This is important for improving the effectiveness of both game features and games as a whole. For instance, what are effective scaffolds in serious games, how to implement the scaffolds, and how the scaffolds affect learning and enjoyment. This requires both feedback from users and assessment of users.

In this regards, several papers have assessed enjoyment factors and motivational outcomes in gaming. Jennett et al. (2008) developed a questionnaire for measuring immersion based on cognitive and emotional involvement, real world dissociation, challenge and control. The measures distinguished between immersive games and non-immersive games. Weibel et al. (2008) investigated the subjective experiences in playing games to find the links between flow, presence and enjoyment for players playing the online game *Neverwinter Nights*. The results indicate flow interposes presence and enjoyment. In terms of the enjoyment factors, Lucas and Sherry (2004) identified six motives of playing computer games: competition, challenge, social interaction, diversion, fantasy, and arousal. Sweetser and Wyeth (2005) established the GameFlow model to evaluate player's enjoyment in games which consists of concentration, skills, control, clear goals, feedback, immersion, and social interaction. Subsequently, Fu et al. (2009) followed up the GameFlow model by constructing a self-assessment of player enjoyment, i.e. EGameFlow questionnaire, with the addition of self-assessment on player's knowledge improvement. Kiili and Lainema (2008) also set up a questionnaire to measure enjoyment based on the flow framework they previously created (Kiili, 2005b). However, the study did not assess the relationship between features in games to flow in general, and learning in particular. This is necessary to optimize features in games that may contribute to both flow and learning, which in turn will either improve the effectiveness of games for learning or aid the selection/improvement of the game services. In other words, this will support educators not only in easy game creation, but also effective game design. Thus, in conjunction with the game services development, we focused on post-hoc evaluation of the game services by answering the following sub-questions: *How do we evaluate game features* *(services) in terms of flow and learning? Is there any relationship between flow and learning?*.

Maintaining flow is also one of important aspects in gaming so that the player engages with games for a longer period. To this end, gaming system needs to actively control the delivery of the challenges in the game, i.e. adaptivity, which is one of important services needed in Serious Games Society. To do so, several models and game AIs were created to represent, for instance, different level of enemies in games (Spronck et al., 2004; Yannakakis and Hallam, 2007), task difficulty in games (Bellotti et al., 2009a), and incongruity-based adaptivity (Van Lankveld et al., 2008, 2010). User performances also can be used as feedback for gaming system to adapt the difficulty level according to, for instance, in-game score or the number of times the player's character is defeated by the enemies. Dynamic adaptivity may also enhance player satisfaction in games (Tan et al., 2011) and could make serious games more efficient learning tools (Van Oostendorp et al., 2014).

To reduce the complexity of embedding adaptivity (or delivery strategy) in games and to assist Serious Games Society with their goal of providing open services for developing serious games[3], we aimed at finding an alternative approach in providing adaptivity in games. This also requires games evaluation in terms of flow and learning as mentioned above. However, the evaluation mentioned above is aimed at post-hoc evaluation, that is, how flow and learning can be measured after the intervention. This approach maybe suitable for ascertaining how flow and learning are related. However, this is probably not suitable for adaptivity since flow and learning at that point is already over. Furthermore, post-hoc evaluation requires a lot of time and effort by the educators. To this end, a growing area includes a branch of neuroscience that is investigating the correlation between user psychological states and the value of physiological signals. Several studies have shown that these measures can provide an indication of player engagement (Janicke and Ellis, 2011; Kivikangas et al., 2011; Nacke et al., 2011). This motivated us to further investigate the use of physiological signals for adaptivity in games.

Several papers have examined the physiological correlates of emotions felt while playing entertainment games. For instance, violent game events evoke Electroencephalographic (EEG) oscillatory, facial electromyography (EMG) of players, or changes in the galvanic skin conductance (GSR) (Salminen and Ravaja, 2008; Ravaja et al., 2008; Ivory and Kalyanaraman, 2007). Thus, the physiological measures have potential for examining a player's flow in gaming and could be useful additional information for adaptivity. In this case, the question in our dissertation is *Can we develop adaptivity as a service by using physiological signals, in particular how flow appears in brainwaves during play, so that real-time evaluation and adaptation may take place?*

[3] http://www.galanoe.eu/,http://www.seriousgamessociety.org/

Given all of questions mentioned above, the subsequent chapter describes the research outline and the sequence of proposed experiments of this dissertation.

Chapter 2

Research outline

"The worst thing a kid can say about homework is that it is too hard. The worst thing a kid can say about a game is it's too easy." - Henry Jenkins

Abstract. In a service oriented architecture (SOA) paradigm, computer applications can be implemented as modular and reusable services. Likewise, games as one of computer applications can benefit from the creation of reusable services for efficient serious games development. This will prevent serious games developers from reinventing the wheel by using the ready-to-use services, and encourage them to publish their own services. To this end, we investigated and implemented several game services for creating serious games, including two of important aspects: game format and adaptivity. This includes the evaluation of games and game features (services) in terms of flow and learning to improve the effectiveness of the services. Furthermore, the relationship between flow and learning is still unclear, although recent studies has indicated a positive correlation between both. Thus, the evaluation of gaming features was intended to clarify the relationship between those two. Maintaining players in flow is also one of essential factors in games which represents delivery strategy (adaptivity) as a service in game development. This can be achieved by balancing the difficulty of the challenges with the players' skill. Several papers have reported various techniques of adaptivity in games. One of promising techniques is the use of physiological signals to infer human emotional states. However, little research has been done into the use of physiological signals for adaptivity in games. Thus, we proposed an experiment to test the use of physiological signals for difficulty adaptation, including real-time adaptivity. We consider this approach as a step into the future adaptivity in games. This chapter outlines the rest of this dissertation.

2.1 Background

To elucidate the outline of the research and our choice for the experiments, we briefly describe the fundamental building block of our research. This started with the motives behind our research as follows.

2.1.1 Game and game features development

First, in the context of technologies in software development, a Service-Oriented Architecture (SOA) can be defined as "a software architecture for building applications that implements business processes or services by using a set of loosely coupled, black-box components orchestrated to deliver a well-defined level of service" (Bloor et al., 2007). SOA is not a technology in itself, but rather a set of ideas, recommendations, policies and practices for architectural design. An SOA approach employs modularization and compositionality to achieve flexibility in the development and to enable the reuse of software parts, in an attempt to manage the complexity of large software systems. Likewise, computer games are basically software systems that may contain many parts. Large scale commercial entertainment games require plenty of resources during design and development with people of various expertises, e.g. artists, character designers, story writers, programmers, art directors, etc.

This also applies to serious games. The difference is that serious games are equipped with serious objectives. Entertainment games have been well established in terms of the producers and consumers shown by many major players in the industries with large market shares. On the contrary, the number of serious games are still limited. Hence, one of the main challenges in the area of serious games is how to enable the proliferation of serious games by making the authoring process easier for developers and pedagogical authors. To tackle this issue, Serious Games Society has recently initiated services for Serious Games that provide serious games developers/researchers with components that can be accessed remotely and can be integrated to a game to provide certain features (Society, 2014). Some available services include commercial licenses, such as scenario branching, user log tracking, and educational data analytics, and non-commercial licenses, such as competence assessment[1]. Our work aimed at extending the services to support efficient serious games development. This includes two of fundamental aspects in game design: the format and the delivery strategy (i.e., adaptivity engine) (Bellotti et al., 2009a, 2010, 2012). Our first study focused on the first aspect (chapter 3 to 5) while our last study focused on the later (chapter 6 to 7).

In designing the game format for educational games, we should also consider flow. To this end, we observed Csikszentmihalyi's eight components of flow that were arranged into the flow framework, i.e. a building block for delivering flow in educational games context (Figure 2.1) (Kiili, 2005b). Flow can be induced in educational games by providing the flow antecendents to the player there consists of clear goals, proper feedback, sense of control, and playability. This is manifested by the game challenges in the form of artefacts that enable fluent use and tasks that encourage cognitive problem solving. During flow, the player fully concentrates in the game, loses self-consciousness, feels a rewarding experience, and loses track of time. The consequences of being in flow

[1] http://services.seriousgamessociety.org/services

are learning and exploratory behavior exhibited by the player. In other words, the framework suggests that properly integrating the flow antecedents into the games (either as game features, game mechanics, or game narrative) will deliver flow to the player, and consequently, enhance learning and exploratory behavior.

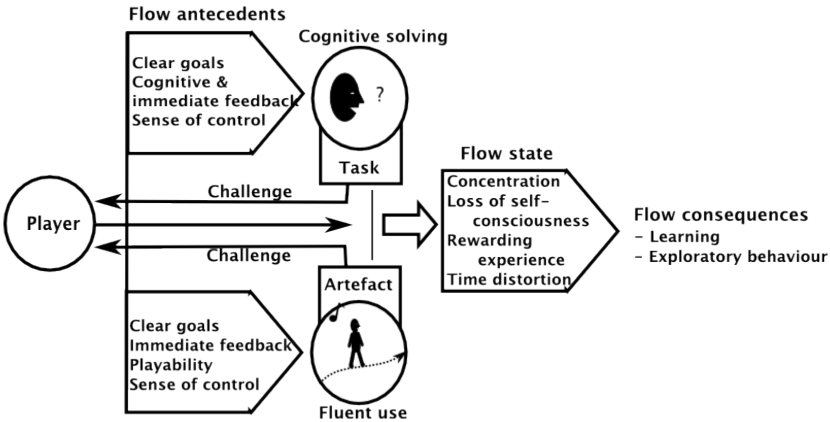


Figure 2.1: The flow framework (adapted from Kiili (2005b)).

Proposed experiments

We created an extensible game format and architecture that supports educators in easy content creation. We created game features using the flow framework and subsequently instantiated a prototype using the format. This would give a better grip on how and if flow in our framework for game development improves learning, and thereby leads to better learning games. We tested the prototype in terms of user performance and user perception to improve the prototype for the subsequent study, i.e., the evaluation of game features in terms of flow and learning. We also set out to determine whether players can benefit from the flow antecedents in the flow framework, and tentatively, which factors contributed the most. However, we should emphasize that the aim of this research is not to validate the framework.

2.1.2 Evaluation of game features

Serious games are different from entertainment games in the sense that serious games should provide 'real-world' educational contents that capitalize on the merits of videogames to offer compelling experiences (flow). In this sense, the players not only should be immersed in the games but also should learn the didactic contents crafted in the gaming environment. Consequently, designing serious games is a non trivial task since the game designers need to balance between learning and enjoyment. Van der Spek (2011) provides several suggestions in designing a serious game without harming the engagement based on the cognitive properties of the human brain, such as the use of visual cues and surprising elements. This can be supplemented to several serious game models for game design and evaluation, such as object oriented design (Game Object Model), experiential oriented paradigm (Experiential Game Model), and learning goal oriented design (Scaffolding Model). However, our research did not address serious game design as a whole, but rather we focused on gaming features (as services). This is essential to improve the effectiveness of game features in games. For instance, as opposed to entertainment games which mostly are constructivist learning tools (i.e., learning by doing), what scaffolds to be used, how to implement the scaffolds, and how the scaffolds affect the learning outcomes and the enjoyment.

Although most of evaluations in game based learning primarily focused on learning performance (Connolly et al., 2012), the engagement was of equal importance. Several works have related the enjoyment in games with learning. Kiili and Lainema (2008) found positive correlation between enjoyment and learning, although the learning was measured subjectively using the flow framework. Likewise, Moreno and Mayer (2007) investigated several design principles for interactive multimodal learning environments which include guided activity, reflection, feedback, control, and pre-training. This may indicate the relationship between flow and improved learning.

To this end, we investigated the use of pseudo tutoring system in games, in particular whether a tutoring tool is effective in educational games without having any detrimental effect on flow. Tutoring systems are associative or task centered which may nurture students in problem solving skills using apprenticeship and problem solving models (Woolf, 2009; De Freitas et al., 2012). For instance, the Andes tutor trained students in solving physics problems (Van Lehn et al., 2005) and improved the average exam score of the students. However, tutoring systems have not been fully exploited in serious games.

Proposed experiments

Using the revised prototype from the previous experiment, we empirically tested two different versions of the prototype in terms of flow and learning. We attempted to determine the effect of a game feature (i.e, a tutoring tool) on players in terms of flow and learning, and clarify the relationship between flow and learning. Furthermore, we aimed at setting up a reproducible procedure to evaluate both flow and learning outcomes in game based learning. This experiment would benefit educators not only in easy game creation but also effective game creation, in particular how a tutoring system (if successful in terms of flow and learning) obviates the need for active guidance and inquiry stimulation by the teacher, thus, making it more efficient.

2.1.3 Adaptivity in games

As mentioned above, one of the keys in efficient serious games development also includes adaptivity. This is important for keeping the player in flow for a longer period by continuously adjusting the difficulty of the challenges in games to match the player's skills. One of approaches is to define factors of flow and create some metrics for each factor. Several psychological studies have attempted to define these factors in games, such as Csikszentmihalyi's eight components of flow, GameFlow model (Sweetser and Wyeth, 2005), Malone's principles of entertainments (Malone, 1981), and Lazzaro's fun clusters (Lazzaro, 2004). In general, the factors include challenge, concentration, player skills, clear goals, feedback, a sense of control, immersion, social interactions, curiosity, and fantasy.

Furthermore, several papers have translated those factors into metrics. For instance, Yannakakis and Hallam (2007) constructed three metrics for challenge in prey-predator games: the average number of steps taken to kill the prey-player over N games, the variance of times taken to kill the player over N games, and the activeness of the predators (opponents) in seeking for the prey-player. The metrics represent control and challenge in games. Subsequently, they combined those metrics into a single model that reflect player's experience in games, and then optimized the model using either simple techniques, e.g. linear regression, or more advanced techniques in artificial intelligence, to match the game challenges and the player's skills. However, a problem with this approach is that the metrics and the model of experience are contingent on the game genres and mechanics. Likewise, Tan et al. (2011) used scores between players and win/lose/draw to indicate gaming proficiency and satisfaction in car race games. Peirce et al. (2008) and Bellotti et al. (2009a) modeled skills and player's competences in form of knowledge space and tree structure, respectively for 3D navigation task based games.

On the other hand, most people share similar physiological traits under certain conditions. Kim et al. (2004) attempted to identify three and four classes of emotion using short-term monitoring of physiological signals with 78.4% and 61.8% of accuracy, respectively. Chanel et al. (2006) used electroencephalography (EEG), skin conductance, blood pressure, abdominal and thoracic movements, and body temperature to identify 3 emotional classes: calm, neutral, exciting. The results show the important of EEG in capturing emotions. Mandryk et al. (2006b) used physiological signals to measure user experience with entertainment technology and subsequently constructed a model for detecting emotion during interaction with play technology (Mandryk and Atkins, 2007). In physical exercise games, Göbel et al. (2010) attempted to use vital parameters to personalize physical exercise games. Salminen and Ravaja (2008) found that electroencephalographic (EEG) oscillatory responses were evoked by violent events in the game. In addition, violent games also affect the facial EMG activity (Ravaja et al., 2008) and skin conductance (Ivory and Kalyanaraman, 2007). Yun et al. (2009) measured facial physiology of the players at a distance for a thermal imaging-based stress monitoring in adjusting the game difficulty levels. This demonstrates the potential of physiological signals for optimizing player's experience in gaming and therefore, we hypothesized that the difficulty adaptation to be generically tractable using physiological signals. In addition, the approach should be feasible for real time adaptation.

Proposed experiments

To provide Serious Games Society with adaptivity, we conducted a study to detect flow using physiological signals, in particular brainwave activity. This would enable the gaming system to adapt its difficulty level according to the player's physiological state. To this end, we customized a game to render three different states: boredom, flow, and anxiety; and we used within-subject design experiments. This approach is one step into the future as commercial products for capturing brainwave activity have also appeared on the market, in particular in the entertainment field (e.g., Emotiv[2], IntendiX[3], Neurosky[4], UncleMilton (Li, 2010),MindGames[5], Mattel[6]), and others are likely to come soon (Nijboer et al., 2011). This experiment would benefit educators in easy and effective game creation since this investigates automatized testing and game balancing to improve learning, so that the educators do not have to determine the proficiency level of every student and balance the game accordingly.

2.2 Outline of the dissertation

We have shortly described the need among Serious Games Society of services for developing serious games. To this end, we approached the need of services in terms of easy authoring/instantiation (i.e., the format and architecture), and flow and learning (i.e., evaluation of game features and delivery strategy). Therefore, this dissertation focuses on those aspects. Firstly, a game format was designed and developed under an SOA platform, and the flow framework was used to define the game features implemented in the platform (Figure 2.2). A game prototype was then developed with a specific, clear, and quantifiable learning goal and evaluated in terms of user perception and performance. This was necessary to improve the prototype for the subsequent phase, i.e. evaluation of services (game features).

[2] http://www.emotiv.com
[3] http://www.intendix.com
[4] http://www.neurosky.com
[5] http://www.mindgames.is
[6] http://www.mindflexgames.com

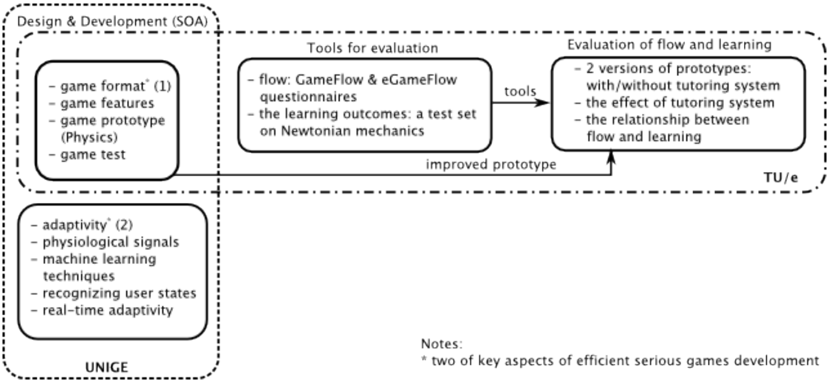


Figure 2.2: The sequence of the experiments

Subsequently, using the game prototype we measured a game feature (tutoring tool) effect on flow and learning. This is essential to find useful game features and their contributions towards flow and learning. Moreover, it serves as a basis for improving the effectiveness of services and games as a whole, or improving the implementation and the selection of services in a game. Thus, assessment tools are needed to evaluate flow and the learning outcomes from playing the games. To this end, we reviewed several questionnaires for evaluating flow and devised a test set to quantify learning outcomes (Figure 2.2). Using the game prototype and the assessment tools, we measured the effect of a game feature (or service) on both flow and learning. We also set out to determine the relationship between flow and learning in games. A study suggested that there was a positive correlation between flow and learning but the study did not quantitatively measure learning (Kiili et al., 2012). In contrast, another study argued that fun learning was ineffective, in particular for promoting deep learning (Graesser et al., 2009). Hence, this experiment is also intended to clarify whether flow and learning can go hand in hand.

Games should also provide balance between the challenges and the player's skills. However, designing a balanced game becomes highly complex as the size of the potential audience grows since different players have different skills and they expect different challenges. In this case, adaptivity mechanisms become necessary to regulate the delivery of challenges. Player and task modeling could be useful for representing the player's ability and the difficulty of challenges for adaptivity, but they are still contingent to the characteristics of the audiences. On the other hand, physiological signals may serve as an alternative or provide additional information for adaptivity since human share similar physiological traits in many circumstances. Furthermore, in contrast to post-hoc evaluation, physiological signals may enable real-time evaluation and adaptivity, by observing how flow appears in physiological signals during play. Thus, we performed experiments on the use of physiological signals to support adaptivity in games (Figure 2.2).

The outline of this dissertation can be explained as follows. To design and evaluate a serious game for learning and flow, in Chapter 3 we designed and developed a game for learning physics that integrates simulation and a tutoring tool using the the flow framework. We hypothesized that all factors in the flow framework were equally important and thus, we had to accommodate all factors into the game prototype. Subsequently, we evaluated the game prototype for usability, usefulness and enjoyment. The results then serve as foundation for further study in evaluating flow and learning.

In Chapter 4, we presented state-of-the-arts in evaluating subjective flow, and we constructed a test set to measure learning outcomes based on the designated learning goals in physics, in particular Newtonian mechanics. In Chapter 5 we revisited our game prototype in the perspective of scaffolding model. We then performed an experiment using two versions of the prototype for learning physics: with a tutor as scaffolding and without a tutor. We hypothesized that the one with a tutor would receive lower level of flow since it might obstruct flow, but it would give better learning outcomes. We used the instruments explained in Chapter 4 to measure flow and learning.

For our work on adaptivity, we briefly explained several physiological signals and their use in research, in particular in the gaming area in Chapter 6. These include electroencephalography (EEG), galvanic skin resistance (GSR), and photoplethysmography (PPG). Following this, in Chapter 7 we performed experiments for adaptivity as described in this chapter. Here, we selected and modified games for the experiments. We then measured the physiological signals and performed exploratory data analysis and predictive analysis on the data. The results of the experiments serve as evidences on the use of physiological signals in adaptivity, both offline and real-time adaptation.

Finally, Chapter 8 ponders upon all the experiments, distills the relationship between the results and the theory, provides a general conclusion to the dissertation, and also discusses the limitations and recommendations for both game developers and future researchers. The dissertation concludes with the summaries in English. Figure 2.3 concisely described the structure of this dissertation.

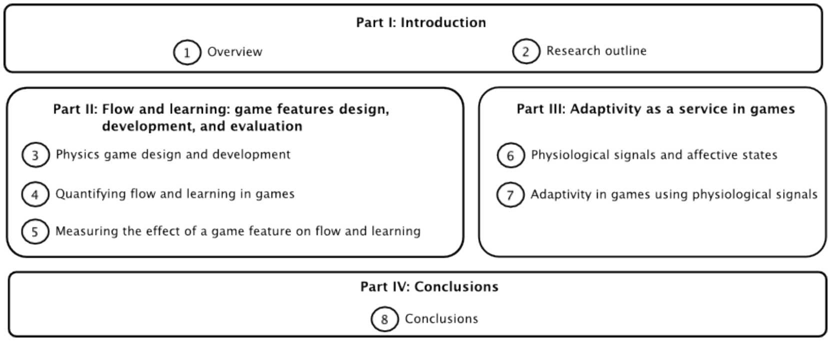
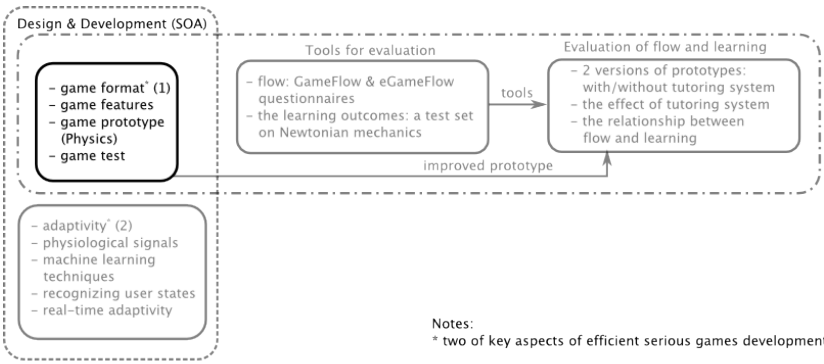


Figure 2.3: The structure of the dissertation

Part II

Flow and Learning: Game Features Design, Development, and Evaluation

Case Study: Physics Game Design and Development



"Games are the most elevated form of investigation." - Alfred Einstein

Abstract. Instruction in physics aims at achieving two goals: the acquisition of a body of knowledge and problem solving skills in physics. This requires students to connect physical phenomena, physics principles, and physics symbols. Computer simulation provides students with graphical models that unite phenomenon and principles in physics. However, such a minimally guided approach may harm learning since it overburdens the working memory. Also, simulation is inadequate in promoting problem solving skills since students need to exercise with a variety of physics problems. Intelligent tutoring systems (ITS), in contrast, can train students in solving physics problems. To get the advantages from both simulation and ITS, we created an online puzzle game in physics that combines simulation and a knowledge based tutor (namely QTut). We used the flow framework as a basis for the game design and addressed three challenges: extensibility, scalability, and reusability in developing our game. We tested the game prototype to study how users perform with the tasks in the game and how users perceive the game prototype. The results show that the users perceived the game to be educative and moderately entertaining. We found two out of four flow antecedents, i.e. sense

of control and goal clarity, of the flow framework to be essential to achieve flow in an educational context. On the other hand, conceptual feedback is useful only if users make mistakes and playability may play a limited role for promoting flow.[1]

3.1 Introduction

In this part, we directed our focus on game format and several features in games for easy and effective serious games development. Furthermore, we evaluated the effect of a game feature on flow and learning. The rest of the chapters in this part (i.e., Part II) discussed the tools to measure flow and learning, and the effect of a game feature on flow and learning, respectively.

To explore certain game features that might contribute to fun and learning in games, we designed and developed an easy-to-author game system. This will benefit the educators in easy game creation. Furthermore, we considered flow in game design and development, in particular how and if flow for game development improves learning, and thereby leads to better learning games. We opted for physics for the game prototype, since physics is fundamental knowledge and a prerequisite studied by students majoring in natural sciences and engineering before they further advance their knowledge at a more specialized level. We selected two concepts in physics for the prototype: force and torque. Both concepts are closely related in which learning the latter requires students to understand the former. The game concept emanated from the learning goals of physics.

The remainder of the chapter is organized as follows. Section 3.2 discusses approaches in learning physics. Section 3.3 and 3.4 explain the game design and development, whereas Section 3.5 explains the user tests. Section 3.6 provides the discussion and conclusion of our paper, followed by subsequent steps in Section 3.7.

3.2 Learning Physics

Instruction in physics aims at achieving two goals: the acquisition of a body of knowledge and the ability to solve quantitative problems in physics. To achieve the learning goals, physics instructions should examine the knowledge structure of physics. In physics, the body of knowledge is organized into three levels: the macroscopic level corresponds to physical objects, their properties and behaviour; the microscopic level explains the macroscopic level using concepts, theories and principles of physics; and the symbolic level represents the concepts of physics as mathematical formulae (Johnstone, 1991). Consequently, physics instructions need to advocate the connection of those levels to the students.

[1] this chapter is based on (Pranantha et al., 2014, 2012a,b)

Lack of knowledge and/or misconceptions at the microscopic level lead students to difficulties in solving physics problems (Heyworth, 1999). The use of concrete models, analogies and graphics may help students to overcome difficulties. Constructivist teaching has the greatest potential to enhance learning where learners actively construct knowledge through inquiry, apprenticeship, and collaboration (Woolf, 2009). In this regard, computer simulations graphically model physical objects and unite the macroscopic, the microscopic, and the symbolic levels. This approach urges students to actively seek questions, explore the simulation, and discover knowledge based on their observations.

However, such a minimally guided approach may harm learning since it does not align with working memory limitations (Kirschner and Clark, 2006). This, to some extent, necessitates the use of scaffolding, which is essential particularly for inquiry learning, a constructivist learning method that actively poses and answers questions to develop knowledge, rather than simply passively receiving established facts (De Jong, 2006). The use of scaffolding reduces the cognitive load of the students when using a computer simulation. A meta-analysis also supports the use of additional instructions in learning with the simulation (Alfieri et al., 2011). Moreover, guided inquiry learning also helps students to plan their simulation experiments (Bonestroo and de Jong, 2012). Traditional instructions enhanced learners understanding of the simulation (Kolloffel and de Jong, 2013) and the use of concept mapping with simulation enhanced deep learning (Gijlers and de Jong, 2013).

Numerous computer simulations for learning physics are available in the market. For instance, the PhET project provides a variety of interactive physics simulations (Perkins et al., 2006). The PhET project investigated several design factors of engaging and effective simulation (Adams et al., 2008a,b). The findings suggest that providing driving questions encourages students to explore the simulation (Adams et al., 2008c).

On the other hand, using simulation alone is insufficient to improve problem solving skills that most students find difficult. It is also crucial for students to practice with a variety of physics problems and to perform retrieval exercise at microscopic and symbolic levels (Karpicke and Blunt, 2011). Associative or task centered approaches, such as Intelligent tutoring systems (ITS) and Fading Worked Example (FWE), nurture students in problem solving skills using apprenticeship and problem solving models (Woolf, 2009; De Freitas et al., 2012). For instance, the Andes tutor trained students in solving physics problems (Van Lehn et al., 2005) and improved the average exam score of the students. Likewise, FWE supports effective learning, but combining ITS and FWE did not contribute to better learning (McLaren et al., 2008).

Combining a physics simulation with a tutoring system may provide students with a graphical tool for exploration (the macroscopic level) and a training tool for problem solving (the microscopic and the symbolic levels). One possible approach is using serious games to combine both simulation and tutoring systems. In fact, games themselves can be seen as a simulation environment with clear objectives to provide a rewarding experience for players. Serious games have the strength of appealing and motivating students (Connolly et al., 2012). Meta analysis also showed that games can be more effective than traditional instructions, but only when considering working memory limitations (Wouters et al., 2013).

To this end, we created an online puzzle game in physics that uses simulation to represent physical objects at the macroscopic level and a knowledge tutor (namely QTut) to explain physical phenomenon at the microscopic and the symbolic levels. Thus, the game graphically simulates the macroscopic level, whereas the knowledge based tutor explains the physical phenomenon at both microscopic and symbolic levels. The game was implemented using HTML5, JavaScript, Box2D-JS, PHP, and Ajax (Asynchronous Javascript and XML) for rich web experiences, JSON (JavaScript Object Notation) for lightweight data storage, and NLTK (natural languange tool kit) for natural language processing[2]. We used rapid prototyping to iteratively create prototypes over short period. We then tested the game prototype with users to collect data on their performances and perceptions of the game.

3.3 Game system design

3.3.1 Designing educational games and flow

Several researchers attempted to integrate pedagogy into the game design process (Quinn, 1994; Amory and Seagram, 2003). However, the works did not integrate these aspects adequately since they did not approach games from an experiential perspective (Kiili et al., 2012). Likewise, designing systems in general has been associated with creating optimal user experience. Therefore, beside the ease to complete a task, user experience also emphasizes the interactions between users and artefacts to accomplish the task and the experience as a result of that context of use. Consequently, user experience is the interplay between three elements, i.e. user, artefact, and task, as shown in Figure 3.1 (Kiili et al., 2012). The characteristics of users, e.g. prior knowledge, determine how users perceive an artefact (usability) and the task at hand (engagement) that can lead to an effective and efficient interactions for completing the task. In addition, an artefact should contain the right functions so that users can perform their tasks efficiently to achieve their goal (usefulness).

[2] http://www.nltk.org/

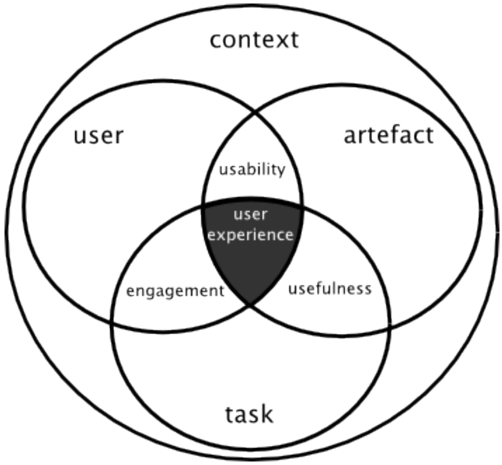


Figure 3.1: The elements of user experience (adapted from Kiili et al. (2012)).

Likewise, in an educational context, designing educational artefacts means being able to understand how users interact with different types of artefacts and how this interaction affects the users' educational experience (Kiili et al., 2012). In this regard, understanding working memory limitations is paramount to ensure that the users are able to master an artefact easily and to learn effectively. However, a learning task should impose a necessary germane cognitive load for knowledge construction so that the users comprehend how to complete the task and why the task is important (Sweller et al., 1998). Furthermore, the learning task should be engaging so that the users willingly spend more effort to complete it. Good usability, useful artefacts and engaging tasks (challenges in games) are prerequisites for a good experience with a learning tool.

Therefore, to improve experience in learning with games, Kiili et al. (2012); Finneran and Zhang (2003) arranged user, artefact, and task elements into a the flow framework for designing flow in educational games (Figure 3.2). There are three phases in the flow framework: a) inducing the flow antecedents, i.e. factors that contribute to flow and should be considered in educational game design, b) achieving the flow state, i.e. an experience where players are completely unaware of their surroundings since they are fully concentrated on solving the tasks in games, and c) obtaining the outcomes of being in flow in gaming (flow consequences) which include learning and exploratory behavior.

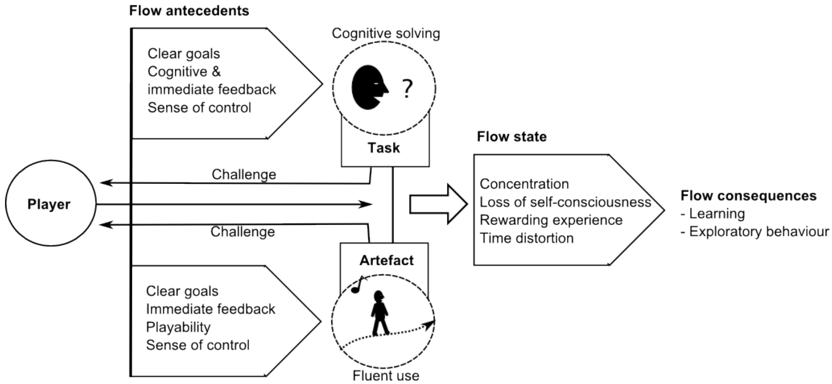


Figure 3.2: The flow framework.

The flow antecedents include clear goals, good cognitive/immediate feedback, and autonomy for performing cognitive tasks while engaging with artefacts (Csikszentmihalyi, 1992), with the addition of playability for the artefacts. Thus, the premise is that games that are well equipped with the antecedents in the form of proper challenges (stimuli) are more likely to promote users reaching the flow state and, subsequently, better learning. This requires proof that we will further explored in Chapter 5. To do so, it is important to integrate the flow antecedents into game mechanics and gameplay. In addition, we need to consider the instruction process - presenting new information, integrating new knowledge, and connecting new knowledge with prior knowledge - to better support learning (Ferguson-hessler and de Jong, 1991). This can be achieved by introducing easy to comprehend game mechanics and gameplay.

3.3.2 Game mechanics, gameplay, and the flow antecedents

There are many definitions of game mechanics in game design. Hunicke et al. (2004) in the MDA (mechanic, dynamic, and aesthetic) framework describe game mechanics as various actions, behaviors and control mechanisms afforded to the player within a game context and together with the game content (levels, assets and so on) to support game dynamics. Sicart (2008) used terminologies in object oriented programming to define game mechanics as methods invoked by agents for interacting with the game world. Thus, game mechanics are a set of rules that bound players in the game world. For instance, the game mechanics of Mario Bros, a well-known Nintendo videogame[3], involve the agent (i.e. Mario) running and jumping in the game world, with the option of power ups that makes the agent either immune to enemies or able to shoot the enemy (Figure 3.3).

[3]http://www.nintendo.com/

On the other hand, gameplay refers to the formalized interaction that occurs when players follow the mechanics (rules) and experience its system through play (Salen and Zimmerman, 2003). Therefore, gameplay dictates how a player interacts with the game to achieve the required goal (winning condition). For instance, in Mario Bros, a player has to complete all levels to achieve the winning state. The player gains points by defeating multiple enemies consecutively and can participate in a bonus round to gain more points. As the game progresses, elements are added to increase the difficulty.

Figure 3.3: Mario Bros video game.[4]

To develop the game mechanics and the gameplay, we considered two educational artefacts in the physics game: a simulation and a tutoring tool (QTut). Tasks in the game include understanding the physics concepts (conceptual knowledge) and solving physics problems (procedural knowledge). To be easily grasped, we selected puzzle solving as the primary mechanic of the game with the tutor as scaffolding. Table 3.1 shows the game mechanics in relation with the flow antecedents.

[4] taken from http://www.digitalspy.co.uk/gaming/levelup/a442090/super-mario-bros-retrospective-platforming-gold-from-the-8-bit-era.html

Table 3.1: Game mechanics and flow antacedents of artefacts and tasks in game for learning physics.

No	Game mechanics	Flow antacedents	Elements
1	A task is defined as a puzzle where the system poses the puzzle and the player solves the puzzle in turn	Clear goal	Task
2	Game level consists of a sequence of puzzles		
3	Game level is either unlocked or locked		
4	Required metrics for unlocking a level are game score and collectibles (e.g. star) in its preceding level		
5	Both the selected and the correct answers are immediately highlighted after a user answering the puzzle	Cognitive feedback	
6	The tutor immediately provides customized text-auditory feedback		
7	Puzzles are given with increasing difficulties in each level	Sense of control	
8	Topics are interrelated for successive game levels		
9	Checkpoints are available in each game level		
10	The tutor may provide hints	Clear goal	Artefact
11	Proper symbols for representing game levels (e.g. grid lock to represent locked levels)		
12	Scaffolding using visual feedback from the simulation	Immediate feedback	
13	Scaffolding using text-auditory responses from the tutor		
14	Grouping functions of game elements into the same grid to ease navigation	Sense of control	
15	Freedom to explore the simulation (to select, to move, to rotate, to collide objects)		
16	Functionality to reset the simulation		
17	Freedom to query the tutor		
18	Providing relevant tools to solve the puzzles if necessary		

Continued on next page

Table 3.1 – continued from previous page

No	Game mechanics	Flow antecedents	Elements
19	The use of simulation to mimic real object behavior	Playability	
20	The use of tutor to mimic teacher		
21	Cartoonish visual graphics for the simulation		
22	Selectable cartoonish avatars for the tutor		
23	Musical background during play		

Using the game mechanics we constructed the gameplay. All game levels are initially locked except at the base level (level 1). For simplicity, all tasks in a level have equal weights for scoring. However, each level has 3 most difficult tasks, each of which is indicated by a star. If a student answers a starred task, he will receive one star.

A level has a topic related to its preceding and succeeding levels. For instance, force and torque can be two successive levels. If a level is unrelated to the preceding one, the tutor presents an introduction to denote a topic transition. A student may progress to a level (i.e., unlock a level) if he has passed its preceding level. A student completes a level if he earns at least two stars and scores above a certain threshold. During the game, a student may query the tutor about concepts, formulas, and terminologies. Moreover, relevant tools, e.g. ruler and calculator, can be used to help solving the puzzles. There is no timeout in the game but we use the timer for logging purpose.

3.4 Game development

To improve applicability and generalizability, we started by identifying challenges in the development of the game, devising the solutions, implementing each solution as a module, and ended with integrating the modules into a complete system. We considered three challenges in developing the game system: extensibility refers to the ease to produce a variety of games for different topics, scalability means the ease to attach new modules to the system, and reusability corresponds to the use of some modules for other purposes. Therefore, to address the challenges, we created a game format and a knowledge based tutor. In addition, we implemented the system in a modular fashion.

3.4.1 Game level and Game format

We used game levels and created a game format to allow extensibility (Prananatha et al., 2012b). The game level clusters learning topics into levels based on their complexity. The game format sets each game level as series of tasks -a puzzle set- drawn from the database (a JSON file). A task - or a task item- is either a closed ended question about a simulated event or an action request in the simulation area.

Figure 3.4 described a puzzle set that consists of several task items. Each task item has two types of data: the scaffolding data and the simulation data. The scaffolding data (Program 1) has an id, a question, a list of feedbacks, a sequence of possible answers, and an index of correct answer. Subsequently, the simulation data (Program 2) includes a collection of objects and a list of available responses to the action request.

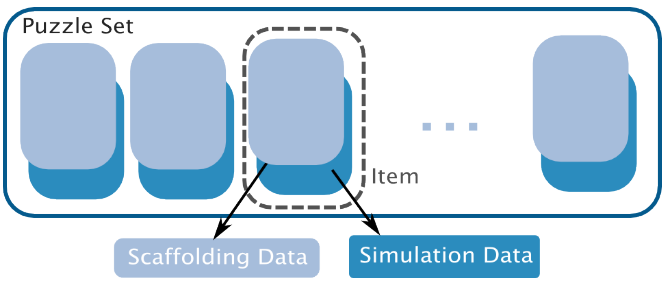


Figure 3.4: A puzzle set and a task item.

```
Program 1 (Scaffolding data):
{
      id  : 1,
      question : "What is the friction force?",
      feedback : ["Friction force = Normal force
                   x cos(alpha), where alpha is the
                   angle of the friction force with
                   respect to horizontal ground",
                  "Well done"],
      answerLst  : ["120 kg.m/s2","115 N","117 N",
                    "114 kg.m/s2"],
      idxCorrectAns: 2
}

Program 2 (Simulation data):
{
      id  : 1,
      objects : [ {
                   "id": 1,
```

```
                    "name": "crate",
                    "class": "crateActor",
                    "position": {"x": 0, "y": 230},
                    "size": {"w": 640, "h": 20},
                    "image":"images/force/crate.png"
                    },
                    ...
                ],
      responseLst: {
                      "response": ["createJoint"],
                      "objects":[{
                                    "to":["extension"],
                                    "anchor":{"x":1,"y":1}
                                }]
                     }
}
```

Using the game format, a game consists of a sequence of inter-related tasks that can be easily created to learn problem solving skills. Some tasks can be recalled several times to promote a retrieval practice, which is essential for learning (Karpicke and Blunt, 2011).

3.4.2 Knowledge based tutor

Beside the scaffolding data in the task item, we created QTut, a knowledge based tutor implemented as a service in Simple Object Access Protocol (SOAP). This enables other games or modules to use QTut by calling the service. QTut allows students to query some information in relation to the task at hand.

To support the extensibility of QTut, we created *knowledge triplet* (Qs, R, DA), where Qs refers to a list of query samples; R represents a response to a list of query samples Qs; and DA denotes a dialog act (Program 3). The knowledge triplet (subsequently called triplet) represents QTUt knowledge on learning topics. Consequently, the number of triplets is contingent on the coverage of the learning topics in the game.

```
Program 3 (A knowledge triplet):
{  "Qs": ["Define normal force","What is
         normal  force"],
   "R": "Normal force  (N) is the component
         (perpendicular to the surface
         of contact) of the contact force
         exerted on an object by,
         for instance, the surface of a
         floor or wall, preventing the object
         from penetrating the surface",
   "DA": { "key": ["what", "define"],
```

```
            "intention": "ASK_EXPLAIN" }
}
```

Using NLTK, we use the triplets to construct an N-gram term frequency - inverse document frequency (TF-IDF) table (Table 3.2) that measures how concentrated the occurrence of a given word is in a collection of triplets. Words with high TF-IDF numbers imply a strong relationship with the triplet they appear in, suggesting that if that word were to appear in a query, the triplet could be of interest to the student.

Table 3.2: An example of N-gram TF-IDF table with 2 triplets.

N-gram words	TF-IDF of triplet 1	TF-IDF of triplet 2
Net force	0.40	0
Normal force	0	0.4
Force	0.10	0.10

TF-IDF is computed as follows. Suppose we have a collection of N triplets. Define f_{ij} to be the frequency (number of occurrences) of term i in triplet j. Then, define TF_{ij} to be f_{ij} normalized by dividing it with the maximum number of occurrences of any term in the same triplet (1) (Rajaraman and Ullman, 2011)).

$$TF_{ij} = \frac{f_{ij}}{\max_k(f_{kj})} \tag{3.1}$$

whereas the IDF for a term is defined as follows. Suppose term i appears in n_i of the N triplets in the collection, then,

$$IDF_i = \log_2 \frac{N}{n_i} \tag{3.2}$$

The TF-IDF score for term i in triplet j is then computed as

$$TF - IDF_{ij} = TF_{ij} \cdot IDFi \tag{3.3}$$

To match a user query to a triplet, we also transformed this query into a set of N-gram words (Figure 3.5). We developed a Naive Bayes classifier to determine the similarity between the set of query words and the triplets using TF-IDF information (Manning et al., 2008). QTut subsequently ranks the similarity values in descending order and removes triplets that have similarity values below a certain threshold. QTut performs intention matching on the dialog act (DA) of the remaining triplets with the following rules: if it finds a match, then it returns the corresponding triplet; otherwise, it returns the triplet with the highest similarity value.

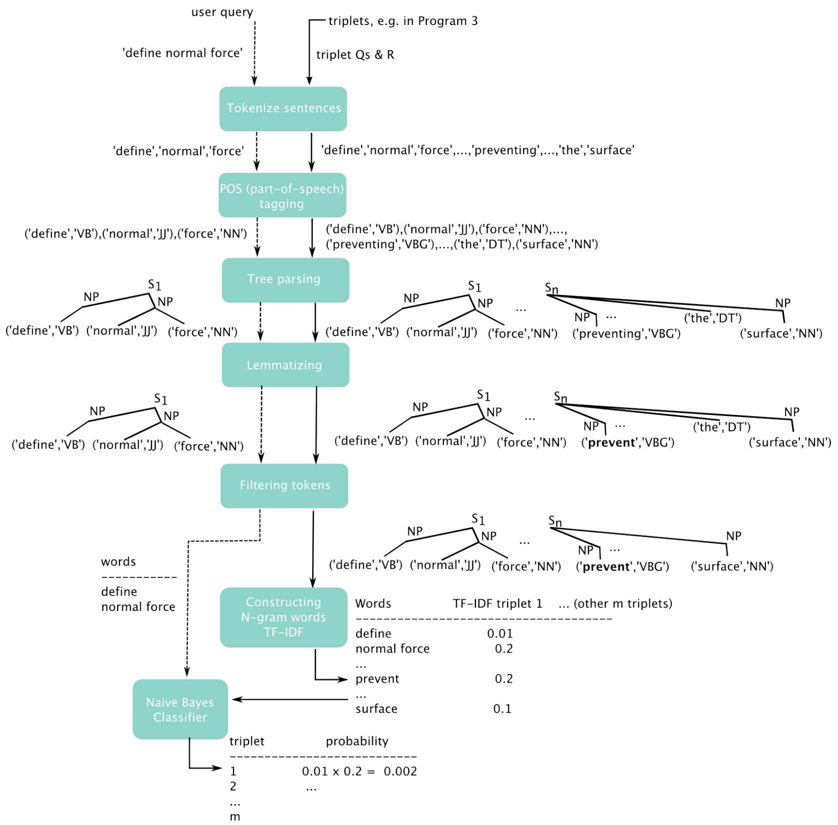


Figure 3.5: Low level natural language processing (NLP) for triplets and a user query.

QTut has two response modes: "text" and "text-auditory". For text-auditory mode, we use a free text-to-speech (TTS) web service[5] to convert texts into speech (Figure 3.6). The procedure is that QTut sends the texts to the TTS web API using HTTP GET and the TTS web API subsequently synthesizes the speeches and sends them to QTut. This supports both extensibility and scalability.

[5] VoiceRSS Text To Speech (http://voicerss.org/)

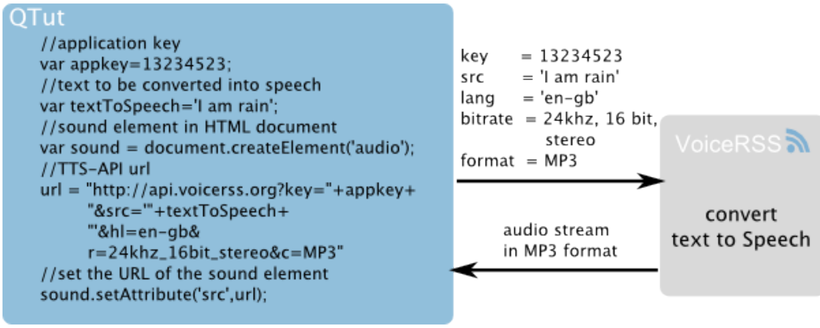


Figure 3.6: Converting Text to Speech.

To visually represent QTut in the game, we created three selectable avatars (Figure 3.7) combined with a user query input field. An avatar, in a broader sense, is an image that represents an agent in interactive exchange, and functions as communication interface linking with the information the user needs (Sheth, 2003).

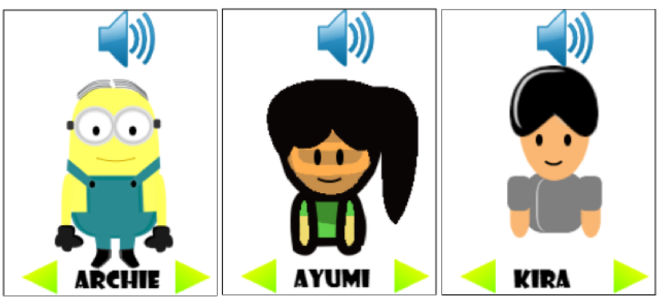


Figure 3.7: Selectable avatars.

One of such famous avatars was Microsoft's maligned help avatar, "Clippit". Clippit is a good avatar: interactive, animated and a reliable source of information. However, the character was too intrusive and distracting people from their work. This differs from modern e-Learning where the user is actively seeking information, knowledge, or skills. Thus, the avatar should not attempt to force information onto the user which confirms our choice to opt for QTut as an inquiry based avatar with limited visual animation.

3.4.3 Modular system

To facilitate scalability and reusability, the game system is divided into functionality modules (Figure 3.8): *a*) tutoring module delivers questions, provides hints and feedbacks, and responds to queries; *b*) the physics simulation module handles all graphical events based on laws in physics; *c*) the delivery module draws a task item from the puzzle set either in random, sequential, or difficulty based order; and *d*) the data module accesses, organizes, and manipulates game database (i.e., game contents, game configuration, and user log).

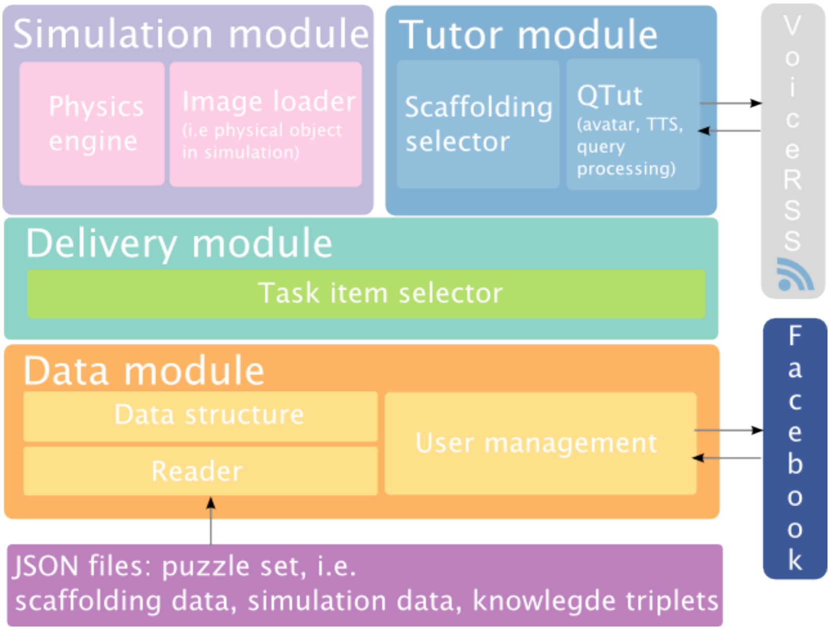


Figure 3.8: A stack of modules as a complete system architecture.

To minimize the needs of user management and to support the game distribution, the system is connected to a social networking platform (Facebook) using *Facebook Javascript API*[6] (RESTful). The system extracts user information on Facebook to be stored into the database.

3.4.4 Graphical user interface (GUI)

Good GUI is essential to improve goal clarity and sense of control of artefacts. To this end, the layout of the game GUI was designed using grid systems to group all elements according to their functionalities. This allows the game users to easily comprehend and navigate the interface (Elam, 2004). Figure 3.9 shows the wireframe of the game GUI: the tutor area on the top right consists of a tutor avatar and an input text to enter a query for the tutor, the information area on

[6] Facebook Developer API (https://developers.facebook.com/)

the middle presents feedback and a task from the tutor, and the simulation area on the bottom plays physics events. The final GUI of the Physics game prototype is shown in Figure 3.10.

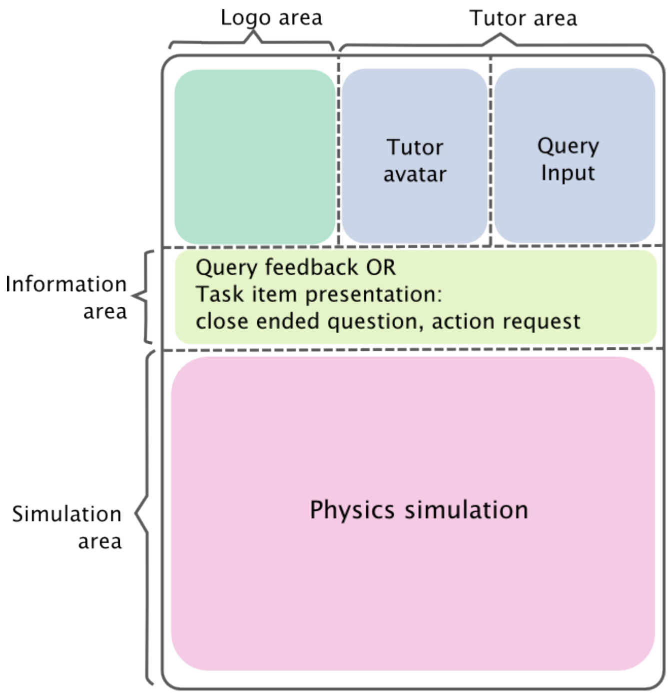


Figure 3.9: the wireframe of the game

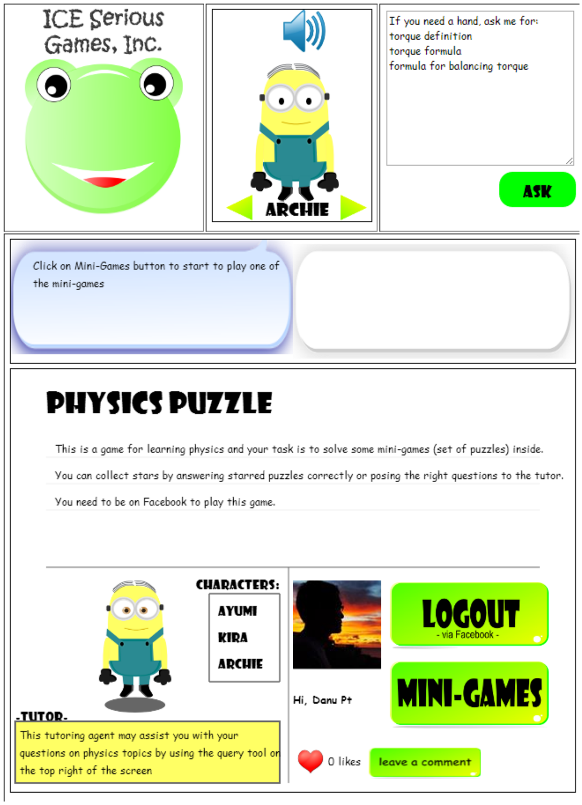
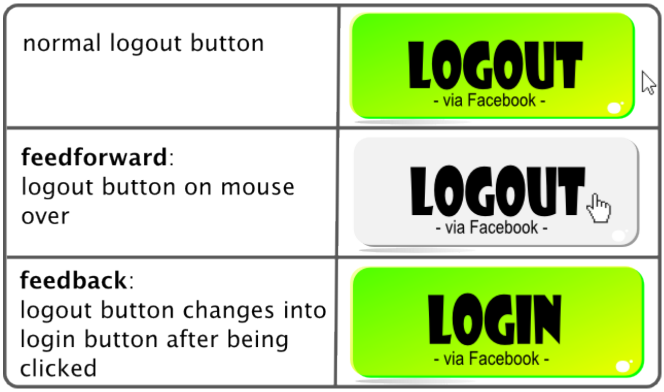


Figure 3.10: the game GUI

The GUI elements (e.g., buttons and playable objects) use the feedforward and feedback concept to allow intuitive interaction. Feedforward is the information that occurs during or after a user action, for instance, highlighted button upon mouse-over. Feedback is the return of information about the result of a process or activity (Wensveen et al., 2004). For instance, clicking on a button opens a new window.

Figure 3.11(a) shows the use of feedforward and feedback in a Logout button and Figure 3.11(b) shows the feedforward for using a ruler in simulation. Feedforward conveys an implicit message that the logout button is click-able by changing its color upon mouse-over event; and the feedback responds to user action (a click) by changing the logout button into a login button. Feedforward is also used to help students in problem solving. For instance, a calculator button appears if a task item asks a student to calculate force. The physics simulation shows a ruler if student needs to measure length or distance.



(a) Feedforward and feedback in a Logout button

(b) Feedforward of a ruler in simulation

Figure 3.11: Examples of feedforward and feedback.

3.4.5 The game prototype

The game prototype was intended for bachelor degree students and it has two levels: force and torque. The first level consists of nine close-ended questions. The questions are either conceptual or procedural problems. The second level has six action requests that demands student to interact with objects in the simulation area. Figure 3.12 shows a list of game levels where all levels are locked except level 1 (force). Figure 3.13 shows a task item in the first level that asks about stationary state. Figure 3.14 shows a task item in the second level that demands students to balance the mobile toy. Each correct answer is awarded with ten points and a star, if the task item is a starred task item. A student passes a level if they earns two stars (three stars are available in each level) and scores above 50% (i.e., 50 points for level 1 and 30 points for level 2).

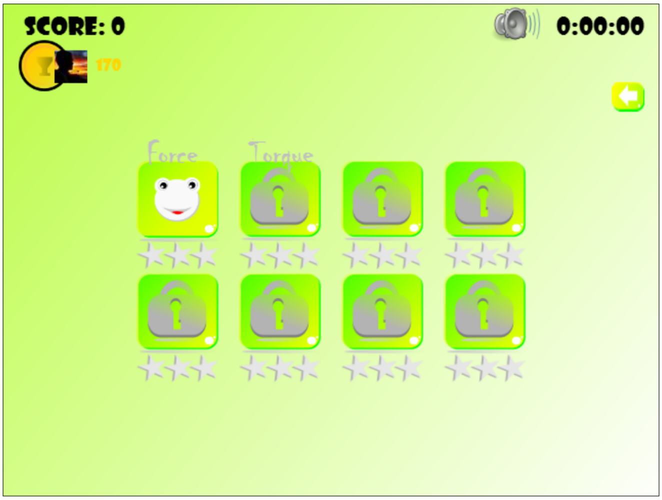


Figure 3.12: Locked game levels

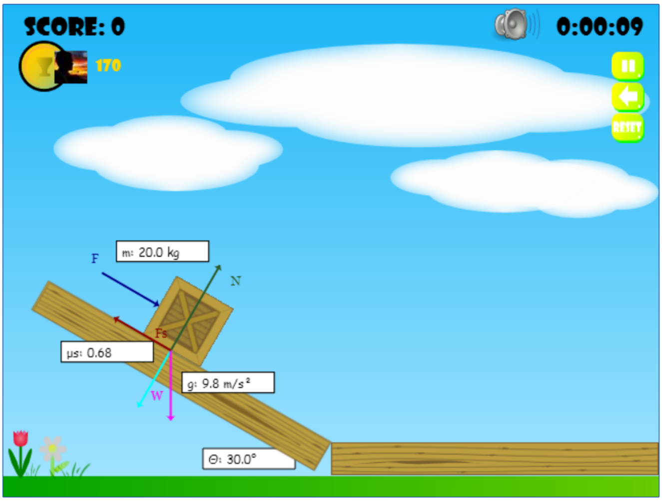


Figure 3.13: Game level 1, force

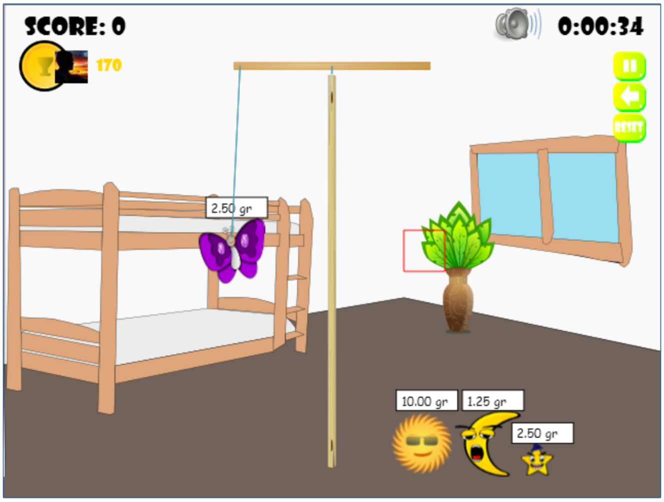


Figure 3.14: Game level 2, torque

3.5 Game evaluation

We evaluated the game prototype to study how users perform with the tasks in the game (user performance) and how users perceive the game prototype (user perception). Here, the user performance reflects usability and usefulness, whereas user perception reflects subjective engagement and usefulness.

3.5.1 Methodology

To test the game prototype, we instructed each user to complete two game levels. Each user had to earn two stars and achieve 50% of points in each game level. The participants might query QTut whenever they needed assistance to solve a task item.

The sequence of the tests can be described as follows.

1. The participants fill out a pre-questionnaire about their knowledge in physics and their exposure to games.

2. The participant plays with the Physics game. Meanwhile, the game system creates three types of logs: the game level log summarizes the progress of the user at each game level, the task log records the user performance in each task, and the tutor log records the dialogs between the user and QTut. In addition, the participants were video logged during playing.

3. The participants fill out a post-questionnaire about their subjective perception of the game, including QTut, the contents, the gameplay, and the enjoyment.

3.5.2 Participants

We recruited 10 participants (graduate and undergraduate students) for the tests (μ = 26.6 y/o, σ = 2.5, 3 participants were female) and each participant was rewarded with 5 Euros.

According to the pre-questionnaire responses, all participants had undergraduate levels of physics or above, except one participant who had a high school level of physics, while classical mechanics (e.g., Newtonian principles) is the most familiar concept.

In daily lives, the participants play games 1-5 times a week (μ = 2, σ = 1) and a playing session lasted for 1 hour on average. Most participants played games for fun and identified themselves as occasional gamers. The pre-questionnaire also showed that a notebook is the most frequent device for gaming among participants. This suits our proposed system well.

3.5.3 Results

We divided the test results into two areas: the user performance based on the game log, and user subjective perception based on the post-questionnaire.

1. User Performance

The participants spent between 14.5 to 29 minutes on completing the game (μ = 19.8, σ = 4.7) (Table 3.3). The mean score was 120 points with a minimum of 90 points and a maximum of 150 points. The final scores of the first and the second participants are missing due to hardware failures during the experiment. There was no significant difference in game time ($F(3, 6) = 0.78$, $p = 0.55$) and final score ($F(3, 4) = 4.81$, $p = 0.08$) between participants with respect to their prior knowledge. All participants retried level 1, whereas 2 participants retried level 2. This was likely because the participants were familiarizing themselves with the games at the first level.

Table 3.3: Users' gaming data.

User	Time (mm:ss)	Score (pts)	Num. of retry level 1	Num. of retry level 2	Prior knowledge
1	18:44	-	1	0	high
2	20:09	-	1	0	medium
3	25:24	90	1	0	low
4	14:56	120	1	1	medium
5	14:28	130	1	0	medium
6	15:33	110	1	0	high
7	14:46	130	1	0	high
8	17:54	130	1	0	very high
9	25:26	150	1	0	very high
10	28:25	120	1	1	medium
μ	19:46	123	1	0.20	-
σ	±4:42	±17.53	0	±0.42	-

Table 3.4 shows all activities performed by the participants. Only one participant used features on Facebook (i.e., like/dislike and comment), whereas most participants did not share their activities on Facebook due to privacy concern. All participants kept the QTut audio on, 5 participants asked some questions to QTut, and 1 participant changed the QTut avatar. Relevant tools (i.e. calculator, ruler, and trigonometry illustration) were frequently accessed by all participants.

Table 3.4: User activities in gaming.

Activity	Num. of performing users
Like/dislike on Facebook	1
Leave a comment on Facebook	1
Turn on/off QTut audio	0
Query to QTut	5
Change QTut avatar	1
Access competition table	1
Use calculator	10
Use ruler	10
Open trigonometry illustration	8

2. User Perception

The post-questionnaire responses showed that the participants found QTut helpful. Nine participants preferred to have QTut synthesized speech since it helped them to learn better, and to retain their attention. This supports the dual channel assumption which was proposed into the cognitive theory of multimedia learning, i.e. the human information-processing system contains an auditory channel and a visual channel (Mayer, 2005). The participants also found QTut responses informative and QTut avatar pleasant (Figure 3.15). Five participants (who made queries to QTut) perceived QTut to be accurate (4.0/5 on a Likert scale).

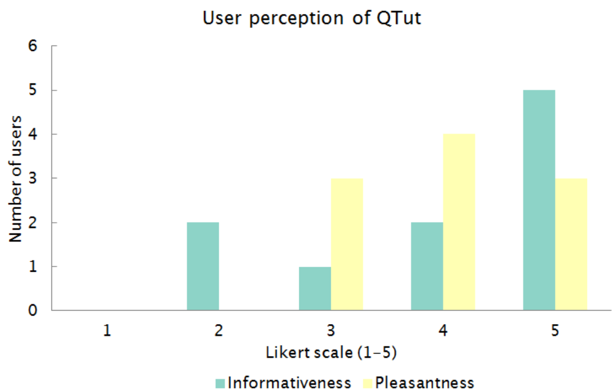


Figure 3.15: User perception of QTut responses and avatar.

All participants agreed that they learned and/or recalled some concepts in physics: force, weight, friction, and torque. Eight participants claimed that they would have comprehended the game mechanics even if we did not provide any instructions (or mission), since the icons and the GUI were very clear. This illustrates the expressive power of the GUI, and feedforward and feedback concepts were successful in delivering such a GUI.

Figure 3.16 shows the participants' perception of the Physics game prototype and Figure 3.17 shows their perception of the game that they favor to play as a baseline. Figure 3.18 compares the average user perception of the game prototype to the game that they favor to play (the baseline). The participants perceived the Physics game prototype as significantly more educational compared with the baseline ($F(1,18) = 22.785$, $p < 0.001$). Although the Physics game prototype was less entertaining compared to control, the difference is not significant ($F(1,18) = 1.056$, $p = 0.318$). In addition, the Physics game prototype offers somewhat equal challenges to the control ($F(1,18) = .051$, $p = 0.824$).

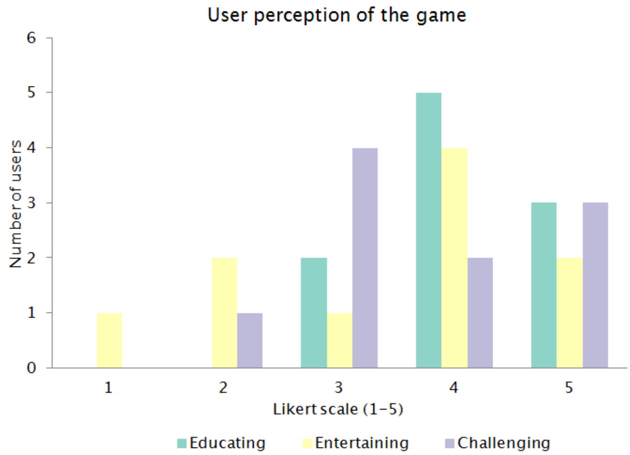


Figure 3.16: User perception of the Physics game.

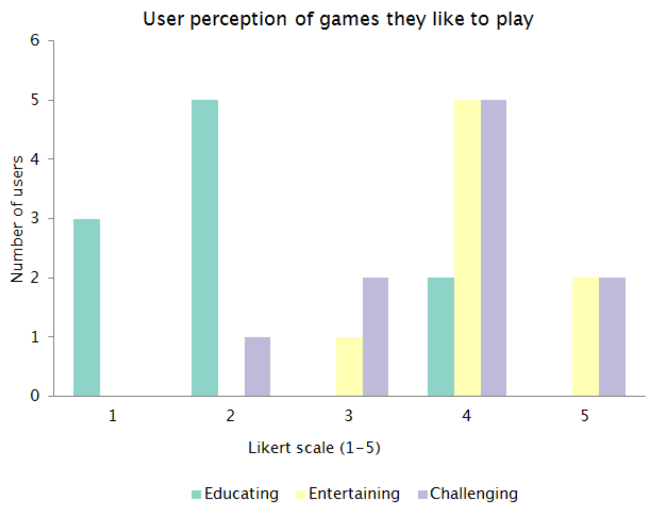


Figure 3.17: User perception of games they like to play (baseline).

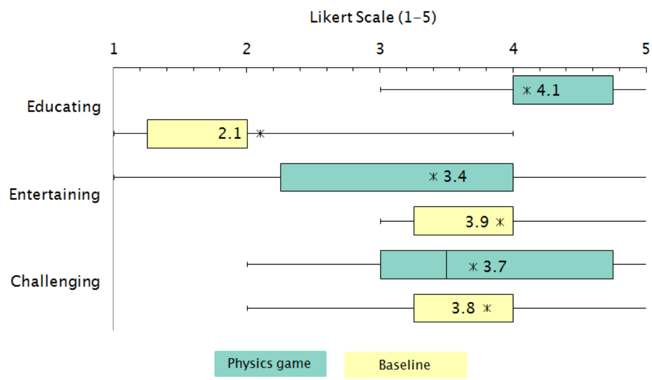


Figure 3.18: User perception of the Physics game against the baseline.

Positive feedback from the participants include good GUI (color and animation), helpful scaffolding tools (feedback and QTut), interactive learning, hard but do-able problems, and challenging game. In addition, the participants preferred some of the game artefacts (i.e. ruler, trigonometry, and calculator) to be always displayed instead of presented only if necessary. Negative feedback mostly involve the learning materials, such as: calculating problems should not be presented successively.

3.6 Discussion and conclusions

In this chapter, we have presented our work on designing a physics game to support inquiry learning and the retrieval practicing using a simulation and a knowledge based tutor (QTut) as services. The game prototype was implemented as an online puzzle game that used driving questions to encourage students to explore the simulation. Online games have become the learning tool that best provides students with enjoyment and increasing sense of immersion (Ampatzoglou and Chatzigeorgiou, 2007; Virvou et al., 2005). We addressed three challenges in designing the system: extensibility, scalability, and reusability. Consequently, we defined the game levels and the game format to cope with extensibility. Also, knowledge triplets were designed to represent QTut knowledge. The system was divided into modules to allow scalability and reusability. The game GUI was designed using feedforward and feedback concepts on a grid system. Subsequently, we tested the game prototype to study the user performance and perception for improving the game.

The results show that users perceived the game as educative and moderately entertaining. The use of scaffolding positively contributed to the game experience. In the game prototype, we selectively presented some of the game artefacts, i.e. the ruler, trigonometry, and the calculator, only if the encountered puzzle required the users to use them. This signaled the users to use the game artefacts for solving the task at hand, and would have reduced the working memory load (just-in-time presentation). However, we found that the users preferred all game artefacts to be always available from the beginning of the game. This might seem contradictory to the working memory limitation. Perhaps, however, displaying all game artefacts from the beginning of the game gives the users a higher sense of control, and affords the users to choose to use them or not. This result is simular to the finding in (Van der Spek, 2011). This tendency seems to agree with the flow antecedents but we have to further investigate its effects on actual learning.

The users perceived the game GUI to be informative, which contributed to the clear goal of the game. Based on our video log, they required 1-2 minutes to get acquainted with the game and subsequently focused on solving the puzzles for the rest of time. All users understood how to query the tutor but only half of the users chose to use the tutor. The users who chose not to use the tutor claimed that they wanted to recall their memories in solving the puzzles on their own. Although this would require longer time, the users felt more challenged. However, we could not determine whether prior knowledge contributed to the choice.

The users stopped for a moment after answering each puzzle to read the conceptual feedback given by the system only if they had made a mistake. They think the feedback is actually important to learn from the puzzle, especially if they make a mistake. On the other hand, the users skipped the conceptual feedback if they had succeeded in solving the task. However, it was unknown whether the users had mastered the task or not. Hence, we argue that a relatively short explanatory feedback or a simple indicator of success may be more effective to create flow if users succeed in a mission, but we cannot corroborate its efficacy for learning. We did not find any user making too many mistakes that might lead to frustration, which means the game difficulty was adequate.

Only one user played with the avatar and no user minded the music background. The users observed and sometimes reset the simulation but they did not play much with the objects within simulation. Some of the users argued that they observed enough information depicted in the simulation at the beginning of each puzzle. Thus, they only paid attention to a specific event in the simulation that related to the task at hand.

We concluded that sense of control and goal clarity are essential in designing flow in educational puzzle games, as highlighted by (Broin, 2011). On the other hand, conceptual feedback is useful only if the user makes a mistake. Although playability was introduced in the flow framework (Kiili et al., 2012) as a new dimension and contribute to flow (Kiili and Lainema, 2008), our finding suggests that the flow antecedents in (Csikszentmihalyi, 1992) to be more valid in our case. This means playability may have a marginal role in delivering flow in learning games. However, this should be treated with care since games require playability to be playful. On the other hand, this could be also related to the game mechanics, i.e. providing a simulation with driving questions might have limited the students' sense of playability.

3.7 Subsequent steps

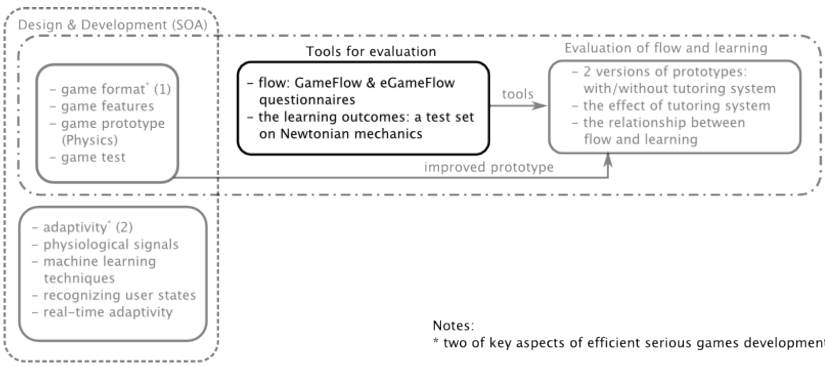
This work provided a baseline for creating educational games and services using the flow framework, in particular supporting the educators in easy and effective game creation. For instance, the use of natural language processing (NLP) in a tutoring system enables the addition of knowledge in the database easier (extensibility). Furthermore, we considered several aspects in flow that may contribute to better learning games (i.e., adaptive feedback, sense of control, and goal clarity). We set the tasks with the increasing difficulties but we have not implemented difficulty adaptation into the system. It will be beneficial to explore the mechanism for difficulty adaptation. Task and user models proposed by (Bellotti et al., 2009a) may fit our case since our game prototype has leveled gameplay. It will also be useful to have more ready-to-use services and a catalog of services grouped by, for instance, functionalities and usage (e.g., services for puzzle game genre). In relation to the flow framework, this also implies the need of a more explicit difficulty adaptation in the framework. Although integrating serious game with Facebook eased the game development process, some participants were concerned about their privacy. Therefore, beside Facebook integration, trustworthy e-learning systems could be used to further ensure that the gaming system is safe. The physics engine (i.e., Box2D JS) enables the creation of a game engine with extensible contents and could be useful for a more complex physics based simulation games.

Our study can be seen as encouraging preliminary results and not as decisive proof of our concept, due to the limited number of participants. Therefore, in-depth investigations into the flow framework as design principles could further refine the framework.

In Chapter 5 we explore the contribution of certain game features in the game prototype to flow and learning. Thus, we perform a more in-depth measurement on enjoyment and learning using the game prototype. The framework for evaluating enjoyment based on the flow theory (Kiili and Lainema, 2008; Fu et al., 2009) is explained in Chapter 4.

Chapter 4

Quantifying flow and learning in games



"That's what games are, in the end. Teachers. Fun is just another word for learning." - Raph Koster

Abstract. Serious games have garnered a lot of attention from teachers and researchers in technology enhanced learning. The pleasant experience induced by playing games has enormous potential to motivate students in prolonging their focus in learning. Consequently, in educational context, flow of using educational games may positively contribute to learning. Having experienced while using a game prototype for learning physics, we need to evaluate the level of flow experienced by players in the game and the knowledge gained by player after playing the game. To this end, we observed two state-of-the-arts questionnaires for measuring flow: GameFlow questionnaire based of the flow framework, and EGameFlow questionnaire based on the GameFlow evaluation model. We compared the indicators and items in each questionnaire and selected the indicators that suitable to our game prototype and our goal. We also created test items to detect misconceptions and to measure knowledge gain in learning physics, in particular classical mechanics. Our approach can be used as a baseline for future research in measuring flow and knowledge improvement in game based learning.

4.1 Introduction

In an educational context, achieving flow in game based learning may contribute to successful learning (Kiili, 2005b). This means users who experience a better flow may learn more compared to user who experience a lesser flow. This is the fundamental building block that underlies the flow framework in Figure 3.2. Intuitively, the design and development of games in an educational context should examine the indicators that contribute to flow stated in the flow framework, as we have demonstrated in Chapter 3.

Having developed the services and the game prototype for learning physics in Chapter 3, the subsequent goal would be evaluating the level of flow experienced by players and the additional knowledge gained by players using the game. This is essential to improve the effectiveness of services (game features) and games as a whole. To this end, we need a mean to gauge flow after playing sessions and the learning outcome between before and after playing the game. E-learning measurement tools missed the fun and challenge in games that make users want to learn (De Freitas and Oliver, 2006; Virvou et al., 2005). To formalize a tool for measuring both flow and learning as a set of key experience indicators (KEI) and key performance indicators (KPI), we need to re-analyze the underlying characteristics of games in general and our game prototype in particular. Moreover, we need to know or establish the learning goal, in particular in basic physics.

Thus, this chapter focuses on means to quantify flow and learning based on state of the arts for measuring flow and learning in educational context. The rest of the chapter is organized as follows. Section 4.2 discusses the quantification of flow and Section 4.3 explains the learning measurement. Section 4.4 provides the discussion on the limitation of the questionnaires for measuring flow and learning and possible future works.

4.2 The indicator of flow

In general, games as entertainment forms have the underlying characteristics of presenting missions to players, providing players with feedback, serving players with sense of victory/loss and immersion, and -in the context of social gaming- allowing players to interact with each other (Prensky, 2002; Rollings and Adams, 2003). Games pose challenges with unpredictable outcome, spur competition between players, and ignite players' curiosity. Several theories have been constructed to evaluate the entertainment aspect of games, two of them are flow theory (see Chapter 2) and Malone's principles (Malone, 1981). Malone (1981) identified four components in which games were able to motivate players: challenge, curiosity, sense of control, and fantasy.

Starting from those theories, several models have been proposed to derive experience indicators that measure enjoyment in games, in particular educational games. Sweetser and Wyeth (2005) proposed a heuristic evaluation of gaming experience that combines usability and user experience in games, namely GameFlow model. GameFlow model includes all indicators (dimensions) that could cause flow stated in the flow theory, i.e. clear goal, autonomy, feedback, concentration, challenge, immersion, and skill, as well as an additional indicator, social interaction. Clear goals (overall goal and intermediate goals), autonomy (player feel a sense of control over their action in the game) and feedback (player receive appropriate feedback at proper time) are interface design heuristics that should allows the player to immediately concentrate on the game. Immersion (deep effortless involvement), challenge, and social interaction are narrative design heuristics to properly influence player's pleasure. Player skills are heuristics that learning design should support player skill development and mastery. Likewise, Kiili (2005b) attempted to establish a framework for measuring the flow state in educational games by defining three phases of flow: the flow antacedents, the flow state, and the flow consequences; along with the heuristic indicators in each phase, such as challenges, feedback, control, and immersion (Figure 3.2).

One of ways to assess each of indicators of flow is by using a questionnaire. A questionnaire is a practical and inexpensive tool to get subjective experience from players themselves which makes it appealing. This led to the development of questionnaires that reflect all heuristic indicators of flow. Kiili (2005b) constructed a questionnaire using 5-point Likert-type response format, namely the GameFlow questionnaire, that represents all heuristics indicators in the flow antecendents, the flow states, and the flow consequences (including learning). GameFlow questionnaire was subsequently tested for its usefulness (Kiili and Lainema, 2008) using RealGame[1], a computer-based interactive business know-how and management training environment - a business simulation game - primary used in business schools and companies. They measured the reliability of each indicator in the questionnaire, except learning, since the learning that took place in RealGame was very difficult to assess. The results showed that challenge and feedback antecedents to be the most important antecedents that supports flow, playability and gamefulness contribute to flow, and the frame story did not correlate to flow most likely due to the insignificant role of the story in RealGame. The GameFlow questionnaire (grouped by indicator) is available in Table A.1 of Appendix A.

Likewise, Fu et al. (2009) reformatted the GameFlow heuristic model (Sweetser and Wyeth, 2005) into a 7-point Likert-type response format questionnaire with the addition of knowledge improvements indicators. This yielded 8 indicators for educational games called EGameFlow with a total of 56 items for all indicators. This include 8 items for capturing player's concentration, 5 items for assessing goal clarity of the game, 6 items for getting feedback quality of the game, 10 items for measuring level of challenges experienced by

[1]http://www.realgame.fi/

player in the game, 9 items for capturing the level of autonomy during play, 7 item for indicating level of immersion, 6 items for representing the aspect of social interaction in the game, and 5 items for assessing subjective knowledge improvement. The questionnaire was then validated using four online games related to learning computer, and statistically analyzed to remove irrelevant items. This resulted in 42 reliable items: concentration (6 items), goal clarity (4 items), feedback (5 items), challenge (6 items), autonomy (3 items), immersion (7 items), social interaction (6 items), and knowledge improvement (5 items). The EGameFlow items are available in Table A.2 of Appendix A.

Since both questionnaires, i.e. GameFlow questionnaire (Kiili, 2006) and EGameFlow questionnaire (Fu et al., 2009), have been previously tested, we chose one of the questionnaires to measure flow in our game prototype in Chapter 3. In this regards, we opted for the EGameFlow questionnaire since it has been used to compare several games. However, we exclude social interactions in the EGameFlow questionnaire because it is not applicable, and added two questions: *a*) Did you experience clear flow and enjoy playing the game? *b*) Did you enjoy the learning process in the game and consider it rewarding?

4.3 Quantifying the learning outcome

Since flow is not about learning, we need to construct a way to measure the learning outcome, which in fact is the main goal in game based learning. Kiili (2006) measured learning subjectively using interview. Likewise, EGameFlow questionnaire marginally measures subjective learning. Thus, we cannot adequately infer the effectiveness of the game -in term of actual knowledge improvement- using EGameFlow questionnaire alone. To quantify the learning outcome, we need to understand the learning goal of the subject at hand so that we can create both the game contents for learning and the test set for evaluation.

Bloom Taxonomy of Educational Objectives, a popular classification of learning goals within education, stratifies cognitive learning into six levels: memorization, comprehension, application, analysis, synthesis, and evaluation (Krathwohl, 2002). Based on this, we focused our learning objectives on the three first levels of Bloom Taxonomy: memorizing and understanding the concepts and formulas in physics, and applying the acquired concepts and formulas to solve physics problems. As mentioned in Chapter 3, a lack of knowledge and/or misconceptions leads students to difficulties in solving physics problems. Therefore, it is essential for the game prototype to support conceptual knowledge acquisition and to reduce misconception. It is also important to address the application of knowledge in physics for problem solving.

To query about users at both the conceptual and the procedural (application of knowledge) levels, we constructed test items that represent conceptual knowledge and misconceptions in physics, as well as procedural knowledge in physics, in particular classical mechanics. The number of test items is limited so that the maximum duration to complete the test is approximately 15 minutes at most.

4.3.1 Conceptual knowledge

Physics is known as a fertile soil for misconceptions among students where classical mechanics has the essential role as the gate to any other topics in physics (Galili, 1995). Classical mechanics, or Newtonian mechanics, concern with the set of principles in physics that describe the motion of bodies under the action of forces. Our game prototype, therefore, introduces and explains main principles in classical mechanics, i.e. Newton's laws of motion including force and torque.

There are various standardized multiple-choice test for assessing knowledge in classical mechanics, such as Force Concept Inventory (FCI) (Hestenes et al., 1992), Mechanic Baseline Test (MBT) (Hestenes and Wells, 1992), and Force and Motion Conceptual Evaluation (FCME) (Thornton and Sokoloff, 1998). Nevertheless, we adapted closed-ended questions used in (Stylos et al., 2010) since they represent simple phenomena in physics that are easily simulated in our game prototype. In addition, the questions may be useful for diagnosing misconceptions among students. The question items are available in Table B.1 of Appendix B.

The first three items correspond to Newton's first laws of motion, i.e. an object either remains at rest or continues to move at a constant velocity, unless acted upon by an external force. In this case, we would like to observe students' understanding of the relationship between force and motion since often students confuse the Newtonian physics with Aristotelian physics, i.e. that the velocity of an object is proportional to the force exerted on the object.

The subsequent two items examine students' understanding of net force and its relationship with mass, i.e. $\sum F = m \cdot a$. In other words, the vector sum of the forces F on an object is equal to the mass m of that object multiplied by the acceleration vector a of the object. One of two items assesses the erroneous perception of an object weight affects the force applied to it instead of the object mass. Item 10-12 further clarify the differences between mass and weight.

Item 6-8 test students' knowledge in action-reaction forces that allows them to decompose force vectors for numerical problem solving (Newton's third laws of motion), and item 9 checks students understanding on the meaning of force. Several articles specifically addressed this issue and they found that most students have poor understanding of both Newton's third law and the concept of force in general (Stylos et al., 2010). The remaining items (13-14) assess the students' understanding of force, torque, and their relationship.

Subsequently, the test results can be interpreted as follows. For instance, if a student answers an item related to the concept of force correctly, he may be able to tell the relationship between force and motion; otherwise, he is most likely to have a misconception. Another example, if a student is able to answer items 10-12 correctly, he is most likely to have mastered the difference and the relationship between mass and weight. Otherwise, his preconception on mass and weight needs to be corrected.

4.3.2 Procedural knowledge

According to Bloom's taxonomy, the learning goals at a procedural level are to use the newly acquired knowledge to solve problems in new situations (Krathwohl, 2002). This requires students to understand the initial and the goal conditions from the problem statement so that they can devise a plan to reach the goal.

In physics, there are two types of strategies in problem solving: working forwards and means-ends analyses (Heyworth, 1999; Larkin et al., 1980; Sweller, 1988). With working forwards, the solver begins with the current information at hand and performs operations in successive order until reaching the goal. Experts exclusively use this strategy to obtain the solution as it is the most efficient approach in particular for familiar problems (Heyworth, 1999). On the other hand, means-ends analyses require forward and backward reasoning that involve a) identifying the goal state, b) finding differences between the goal state and the current state, c) finding an operation to minimize the differences, d) performing the operation, e) repeating steps b to d until a solution is found. This strategy loads working memory with additional temporary sub-goals in reverse order to obtain the solution and is commonly associated with novice problem solving ranging from primary school pupils to university students (Heyworth, 1999). Hence, it is less efficient compared to working forwards strategy. This also means means-ends analysis strategies may inhibit learners in achieving flow.

To measure the procedural knowledge, i.e. the application of Newton's law for calculating net force and net torque, we constructed two types of questions: closed and open ended questions available in Table B.2 of Appendix B. The closed-ended questions emphasis on the end results (i.e. the numerical solutions), whereas the open-ended questions explore the problem solving strategy used by the students.

4.4 Limitations, Possible Research Directions, and the Next Step

The level of enjoyment in games determines whether the player will become involved and continue to learn through the game. This means games should be able to boost learners' self motivation to invest their time in learning and consequently to achieve learning goals. Questionnaires are efficient and effective means to evaluate the subjective enjoyment experienced by learners, while identical pre-test and post-test are common to capture the increasing knowledge of learners such as in (Aleven and Koedinger, 2002; Vosniadou et al., 2001).

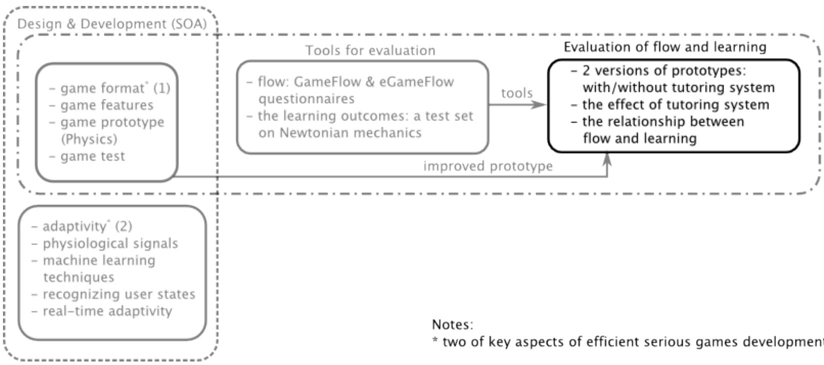
GameFlow and EGameFlow questionaires are somewhat similar and have been shown to have practical merits in measuring enjoyment. The GameFlow questionnaire is a 5-points Likert scale, whereas EGameFlow questionnaires has a 7-points Likert scale. The scale choice is important since it may affect both user subjective evaluation and our interpretation on the user evaluation. It would be interesting to evaluate the choice of scale in such questionnaires. It is worth to note that both questionnaires have not been tested for ceiling effects with highly enjoyable educational games. This is likely due to the limited choices of available educational games in the market. To this end, popular games can be used as benchmarks for a ceiling effect. Previous studies in (Kiili, 2006; Fu et al., 2009) also did not rank the indicators in term of their significance to enjoyment in the context of educational games. This requires the questionnaires to be tested with numerous educational games in various types.

On the other hand, a test set for assessing learning outcome were constructed by assuming our target users have somewhat equal level of knowledge. Consequently, we only put three slightly difficult items into the questionnaire to handle a ceiling effect. If a student succeeds in answering all the questions in the test correctly, we could not infer whether this reflects the student's best ability or the student has far greater knowledge beyond the test set. Time limit for taking the test is then necessary to give a sense of urgency and pressure to the students. An educational research in most common misconceptions in physics can be further explored to enrich the test and to guide teachers in designing test items.

In the subsequent chapter, we used the selected indicators and items of EGameFlow questionnaire and the test set to evaluate two versions of our game prototype developed in Chapter 3: with a tutor and without a tutor, in terms of flow and learning, respectively.

CHAPTER 5

MEASURING THE EFFECT OF A GAME FEATURE ON FLOW AND LEARNING



"No pains, no gains". "If little labour, little are our gains: Man's fate is according to his pains." - Hesperides 752

Abstract. In serious games, learners need to actively construct knowledge by overcoming challenges. This constructivist approach may harm the working memory. According to Zone of Proximal Development (ZPD), learners are capable to learn effectively only under guidance or collaboration with more capable peers. Hence, serious games require scaffolding mechanisms to stimulate learning and guide learners in solving the tasks at hand. Our game prototype for learning physics implemented several scaffolds: hints and feedback, and the main scaffold -a pseudo tutoring tool. In this chapter, we investigated the effect of the tutoring tool on flow and learning in games. To this end, we prepared two gaming conditions: with a tutoring tool and without a tutoring tool, and investigated the difference between both conditions in terms of flow, the learning outcomes (knowledge improvement and misconception), and subjective learning. The results show that the two gaming conditions have significantly different flow - the game with the tutor received a higher score- but no significant difference on the learning outcomes. Both conditions

minimally decreased the misconception among students. The games performed better for improving procedural knowledge with marginal improvement. In terms of subjective learning, we found that the condition with higher level of flow exhibits better feeling of learning, although the difference is insignificant.

5.1 Introduction

Serious games embrace a constructivist approach of learning in which learners need to actively construct knowledge by overcoming challenges in the game, exploring the game mechanics, and following the game narrative. This is in contrast to behaviourist approach that considers learners as passive recipients of information. However, constructivist approaches tend to burden the working memory since learners receive minimal guidance. As mentioned before (Chapter 1) in the ZPD perspective, games may provide learners with guidances. In this light, serious games require scaffolding mechanisms to stimulate learning and guide learners in solving the game tasks at hand. A scaffold is basically a transient entity that assists learner to reach his potential and subsequently removed if learner demonstrate their progress in learning (Lajoie, 2005).

Therefore, in Chapter 3 we have designed and developed our game prototype for learning physics, in which the scaffolds were implemented as hints and feedback, and the main scaffold -a pseudo tutoring tool. Subsequently, in Chapter 4 we explored tools for evaluating our game prototype in terms of flow and learning. In this chapter, we prepared two versions of prototype developed in Chapter 3: with a tutoring tool and without a tutoring tool, to test the effect of tutoring tool on flow and learning. To quantify flow and learning in both conditions, we used EGameFlow questionnaire and a test set described in Chapter 4. Our hypothesis is that the one with the tutor would have a lower flow since it may disrupt the game's continuity.

Previous research showed that there is a loose positive connection between flow and learning in games (Kiili, 2006). Although the learning outcomes were not directly measured, but were examined by interviewing participants (subjective feeling of learning), this provide us with a preliminary hypothesis in investigating the effect of flow on learning. In that sense, we hypothesized that if both game prototypes significantly indicate a different level of flow, then they will significantly affect the learning outcomes. However, we should also stress that this work is about evaluating the tutor, not about flow as a whole. This evaluation will benefit the educators with effective game creation, in particular how a tutoring system as a feature/service (if successful in terms of flow and learning) obviates the need for active guidance and inquiry stimulation by the teacher, and thus, making it more efficient.

Another aspect of learning that need to be addressed is whether games are able to remedy students' misconceptions in the selected subject, i.e. classical mechanics. In the constructivist perspective, the continuous development of human thought relies on a dialectic between experience and concept, and between reflection and action form (Piaget, 1977). In other words, the key of conceptual learning lies in the interaction of the accommodation of schemata to experience (imitation) and the assimilation of experience into existing schemata (play). Nonetheless, learning from experience may undermine misconceptions (Kolb et al., 1984). For instance, students correctly perceived that force may cause an object to move (imitation), but then incorrectly inferred that an object in motion always has force acted upon it (play). This means learning from experience that always see causal phenomena as a whole may lead to misconceptions, which is commonly known as experiential gestalt of causation (Andersson, 1986). Hence, we designed some problems within both game prototypes to address misconceptions and tested whether the main scaffolds (i.e., the tutoring tool) could significantly reduce misconception.

The rest of the chapter is organized as follows. Section 5.2 discusses various frameworks of serious games and scaffolding. Section 5.3 explains the experiment which includes the participants, materials, procedures, and Section 5.4 shows the results. Section 5.5 provides the discussion on the results and the limitation of the study followed by possible future works in Section 5.6.

5.2 Serious games and scaffolding

5.2.1 Models for serious games

Various models have been proposed to infuse learning process into games, among them are Game Object Model (GOM) (Amory and Seagram, 2003; Amory, 2007) and Experiential Gaming Model (Kiili, 2005a). The Game Object Model (GOM) attempts to map pedagogical dimensions and game elements by using abstract and concrete interfaces to represent the pedagogical and design elements, respectively (Amory and Seagram, 2003; Amory, 2007). Kiili (2005a) proposed experiential gaming model based on experiential learning (Kolb et al., 1984) and constructivism. Experiential gaming model describes learning as a cyclic process through direct experience in the game world and learning is defined as a construction of cognitive structures through action in the game world (Kiili, 2005a). The learning processes move continually from active experimentation to reflective observation, to schemata construction, and finally return to active experimentation (Figure 5.1). Other models include Input-Process-Output model (Garris et al., 2002), RETAIN model (Gunter et al., 2006), and 6 I's model (Annetta, 2010).

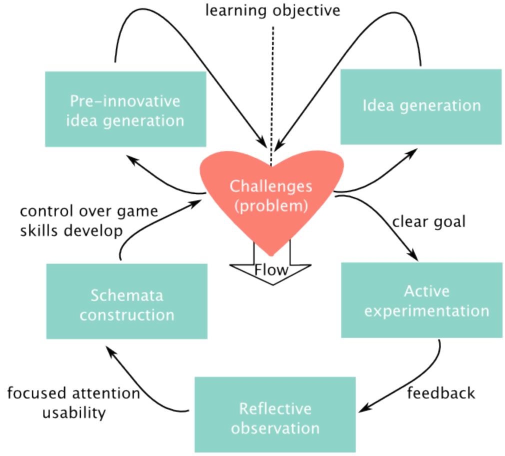


Figure 5.1: The experiential gaming model.

However, having those models, scaffolding in games has been relatively under-studied (Obikwelu et al., 2013). Therefore, they proposed serious games scaffolding model (SGSM) to address this issue. Our approach, however, was based on the flow framework and the games were implemented as problem-solving game. Thus, the scaffolds were infused into the game as part of feedback (Figure 3.2). This approach is somewhat closer to the Experiential Gaming Model (EGM) which was evaluated in similar fashion, i.e. using problem-solving game (Kiili, 2006). The evaluation resulted in an apparent connection between goal clarity and feedback to flow. To understand how scaffolding works in our games, we revisited our games in the light of SGSM.

5.2.2 Scaffolding in our game prototypes

The scaffolds are structured in such a way to keep learners focused on the learning goal. In serious games, the scaffolds include quality feedbacks and hints, which are generally essential to improve learning processes and outcomes (Feeney, 2007). Typically, we can categorize feedbacks in games into three groups: *a*) formative feedbacks are real-time information communicated to the learners as consequences of their actions in games that is intended to modify learners' thinking/behavior to improve learning (Shute, 2008), *b*) summative feedback is a delayed response in the form of performance overview so that learners can reflect and improve on their decisions in games (Gunter et al., 2006), *c*) hints are pointers that guide learners in solving complex problem at hand. The structure of the feedbacks and hints determine their effectiveness in games.

In the SGSM, the scaffolds are applied in games in conjunction with teacher debriefing (Obikwelu et al., 2013). This is to ensure that learners are able to refocus towards the learning goal. Reviewing our games in SGSM, Figure 5.2 shows the structure of the scaffolds in the game prototype with the tutoring tool. The structure is similar for the prototype without the tutoring tool, except that the ask tutor (gray colored boxes) is eliminated. In this model, we started with a brief introduction to the mission in the game. Subsequently, learners, with their initial competence, engage in the problem-solving game with increasing difficulties. In each challenge, learners are provided with hints, a driving question, a simulated phenomena in physics, ask tutor (for the prototype with the tutor), and formative feedback. Summative feedback is then provided in each game level -a subgoal that covers specific topic, e.g. force- for reflection. For instance, if a player incorrectly answers all items related to Newton's first law, the summative feedback will recommend him to be more attentive on this subject. If a learner does not pass a level, i.e. his score is below certain threshold, the learner needs to redo the level with their altered competence, otherwise he cannot proceed to the next level.

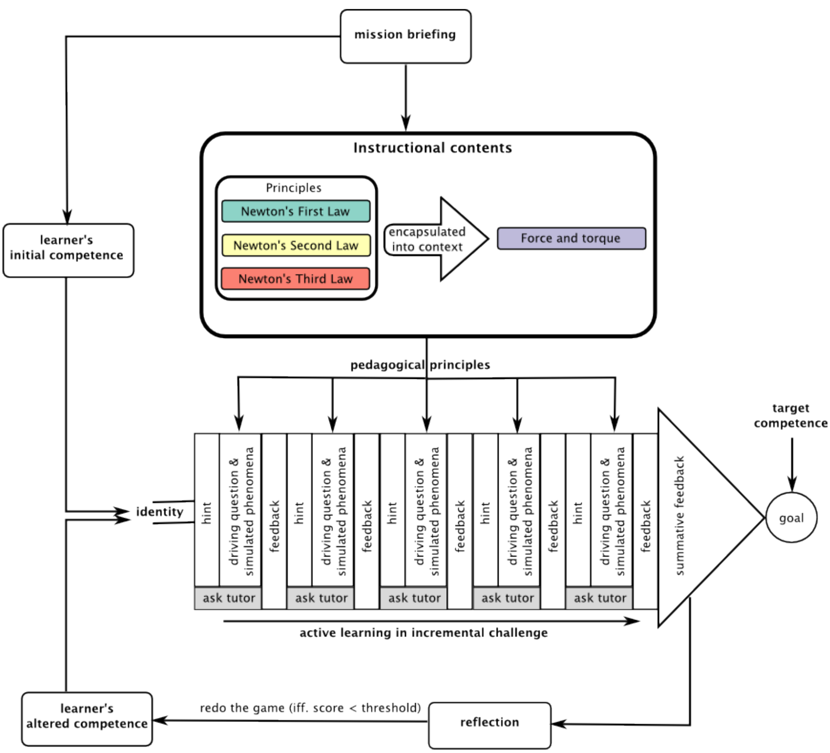


Figure 5.2: The serious game scaffolding model for the game for learning physics.

Each problem in the game has two types of feedback: feedback for correct response and feedback for incorrect response. Feedback for correct response is short without explaining in details the presented phenomena. On the other hand, feedback for incorrect response is longer since it explains the presented phenomena and its relation to Newton's law. Thus, this type of feedback operates as a corrective measure. The tutoring tool enables learners to further query various related concepts and formulas regarding the problem at hand prior to answering it in the game.

5.3 Experiments

In this experiment, we tested two game prototypes and investigated the effect of the main scaffold, i.e. the tutoring tool, on flow and the learning outcomes. Our hypothesis is that the one with the tutor would have a lower flow since it may disrupt the game's continuity, but it would allow a better learning transfer.

5.3.1 Participants

A total of 40 persons: 38 first year and 2 second year undergraduate students in physics department participated in this experiment. However, 2 were discarded after the experiment: one due to missing the EGameFlow questionnaire, and the other one due to missing the post-test. Therefore, valid N is 38. Mean age of the remainder was 19.76 (σ = 1.69), and of these, 16 were female and 22 were male. We asked the participants to rate their prior game experience by letting them choose one of four options to describe themselves: "I hardly ever play games", "I occasionally play games", "I play games regularly" and "I'm an avid gamer". Only 3 out of 38 participants considered themselves an avid gamer, and 6 out of 37 participants considered of themselves to "hardly ever play games" which are too small numbers to measure any effects. Consequently, we opted to combine the first and the second into a group, and the third and the fourth into another group to note their prior experience with games. Casual games and role playing games are the most popular among participants; laptop (personal computer) and smartphone are common devices for playing games. All the participants participated in basic physics course and they claimed to have some knowledge on force and motion in Newtonian mechanics. Each participant was rewarded with 5 Euro for participating in the experiment. A breakdown of the makeup of the groups can be seen in Table 5.1.

Table 5.1: Detail of the participants per condition.

	game with tutor	game without tutor
No of participants	21	17
Male - Female	10 - 11	12 - 5
Game experience[a]	12 - 9	10 - 7
Experience with simulation[b]	5 - 16	5 - 12

[a] *I hardly play games & I occasionally play games - I regularly play games & I am an avid gamer*
[b] *Yes - No*

5.3.2 Conditions

We conducted a between subject design experiment which consisted of two different conditions: a group that allowed to ask the tutoring tool, and a group without the tutoring tool (Figure 5.3). For the tutoring tool condition, we chose procedure that requires learner to consult the tutor. In this case, we provided several topics that can be queried to the tutor that responds in auditory and text mode. In both conditions, we let object of interest in the simulation to glow if it is being pointed. This indicated the object is playable for exploration. Both conditions had background music to accompany learners in playing. This background music could be turned off if learners felt disturbed with the sound.

(a) With tutor (b) Without tutor

Figure 5.3: Two conditions for the experiments.

5.3.3 Materials

We used a subset of EGameFlow questionaire to measure flow (Table A.2 of Appendix A) and two different tests to measure learning: pre-test/post-test for conceptual learning, consist of 14 items of multiple choice questions (Table B.1 of Appendix B); pre-test/post-test for procedural learning, consisting of 8 items (Table B.2 of Appendix B). Besides, we recorded the in-game score of the participants and participants' activities in the game.

5.3.4 Apparatus

The game prototypes were made with HTML5, JavaScript, PHP, Python and NLTK[1], and Box2D in JS[2], a free open source 2-dimensional physics simulator engine ported from C++ to JavaScript. The game contents could be easily created in JSON file format. The game prototypes ran smoothly on two web browsers: Google Chrome and Mozilla Firefox. The experiment was conducted in a computer laboratory room to ensure a minimum of environmental distraction.

[1] http://www.nltk.org
[2] http://box2d-js.sourceforge.net/

5.3.5 Procedure

The participants were sat behind computers in laboratory room and started by filling in a short demographics questionnaire (5 minutes). Subsequently, the participants completed pre-tests: *1*) the conceptual knowledge test (8 minutes), *2*) the procedural knowledge test (7 minutes). After this they were allowed to play the game. The games were timed where the participants had to finish two game levels (level 1: force, level 2: torque) within 20 minutes. The games were easily to get acquainted with regardless the gaming experience of the participants; and all participants completed the game levels. The numbers of completed problems and the total number of problems in each level were also shown in the games so that the players could track their own progress. Each correctly solved problem was awarded wih 10 points. In each level, three featured problems were presented to the players. Each of these problems had a 10 points reward and a star badge if the players correctly solved it. The total score and how many times players redo each level would give us an accurate depiction of how well the players performed in the game. After completing the game, the participants immediately had to fill in flow questionnaire (a subset of EGameFlow questionnaire). Subsequently, the same test procedure was conducted as before the game: *1*) the conceptual knowledge test (8 minutes), *2*) the procedural knowledge test (7 minutes) . The order of the questions were shuffled in the post-test questionnaires. An overview of the procedure is shown in Figure 5.4.

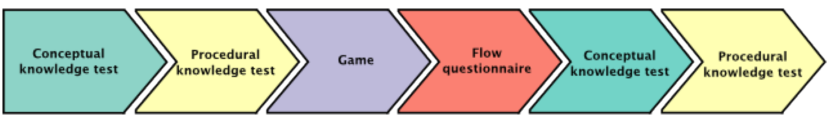


Figure 5.4: Procedure.

5.4 Results

5.4.1 Learning

A. Measured learning outcome

The means and standard deviations of the pre-test and post-test knowledge test (multiple choice only) can be seen in Table 5.2. Both conditions slightly improved the participants' conceptual knowledge although the improvements were not significant, with tutor condition: $t(20) = -1.07$, $p = 0.3$; and without tutor condition: $t(16) = -1.19$, $p = 0.25$. Likewise, the procedural knowledge assessment showed marginal improvements of the post-test over the pre-test, with tutor condition: $t(20) = -1.98$, $p = 0.06$; and without tutor condition: $t(16) = -1.51$, $p = 0.15$.

A full factorial ANCOVA with conceptual knowledge post-test as dependent variable, conceptual knowledge pre-test as covariate and conditions (with tutor vs. without tutor) as factor, shows no significant effect with $F(1,37) = 0.47$, $p = 0.49$. For the procedural knowledge assessment, a full factorial ANCOVA with procedural knowledge assessment post-test as dependent variable, pre-test as covariate and conditions as factor also shows no significant effects with $F(1,37) = 0.01$, $p < 1$.

Table 5.2: Means and standard deviations of pre-tests and post-tests on conceptual and procedural knowledge (multiple choice only).

	Condition	Pre-test	Post-test
Conceptual knowledge	with tutor	7.00(±2.51)	7.38(±2.48)
	without tutor	8.05(±3.07)	8.53(±2.70)
Procedural knowledge	with tutor	1.85(±1.01)	2.43(±1.43)
	without tutor	2.41(±1.46)	2.82(±1.33)

When the experiment was set up, we were expecting to see the additional on-demand scaffolding (the tutor) would provide higher impacts for the players in learning Newtonian mechanics compared to other condition, i.e. without tutor conditions. However, from our experiment, it can be concluded that the additional scaffolding does not suffice in helping the players to outperform other players in the other condition. The additional scaffoldings were more effective for learning procedural knowledge with marginal improvement over the other condition. In addition, both conditions improved the procedural knowledge better compared to the conceptual knowledge. The reason might be that the procedural cues are easier to integrate in games, in particular as corrective measures. This highlights the potential of serious games for training problem solving skills. The positive learning outcomes for both conditions favor games as learning assistance and support the long list of positive effects of games in the literatures (Connolly et al., 2012).

B. Misconceptions

The questions for measuring conceptual knowledge aimed at not only measuring the learning outcome but also to capture the treatment effect on misconceptions. Table 5.3 shows common misconceptions among participants in comprehending Newton's principles for force and torque, mass, and weight. Question items for Newton's first law reveal the perception of the participants concerning the role of force acting on a body when it is moving on constant velocity or is at rest. The results of item 1 show that the majority of participants in both conditions recognized correctly that a body is at rest if the net force acting on it equals zero. However, the pre-test in item 2 shows that most participants in both conditions were thinking in Aristotelian physics, i.e. for a body moves with constant velocity require a constant acting force on the direction of the motion, with tutor condition: 38.10% correct response; and without tutor condition: 41.18% correct response. Both gaming conditions were able to increase the fraction of correct responses by 1.5, with tutor condition: 57.10% correct response; and without tutor condition: 64.18% correct response. In conjunction to item 2, results of item 3 shows that 62% and 47% of participants in with tutor condition and without tutor condition, respectively, preserved the misconception of the Newton's first law. Intervention using game with tutor was able to decrease the false response to 52%, but game without tutor failed to do so.

Question items for Newton's second law capture the understanding of the participants concerning the relationship between mass and total force acting on a body. The results of item 4 show initially 71% participants in with tutor condition and 59% participants in without tutor condition understood that the increasing mass de-accelerates the motion of a body. Both conditions were able to improve the number of correct responses, in particular without tutor condition which significantly improved to 82%. Subsequently, the pre-test responses of participants for item 5 were alarming with only 14% and 23% correct responses for with tutor condition and without tutor condition, respectively. The participants seemed to misidentify mass to weight that affects acting force. The interventions have little effect on improving the results in both conditions.

The results of item 6 show that the participants had rather good application of Newton's third law in both conditions (above 60% of correct responses). However, the intervention in both conditions contributed negatively to their understanding. Results of item 7 and 8 show that the participants had difficulties in applying Newton's third law in real situation. Similar to item 7, the intervention in both conditions had zero to negative effects on the participants' understanding. The pre-test results of item 9 show that most of the participants correlated a body motion with acting force in the direction of the motion, 38.1% in with tutor condition and 47% in without tutor condition. The game without tutor suffered greatly as shown by the decreasing number of correct responses after intervention from 47% to 23.5%, while the game with tutor had zero effect on the correct responses.

On the item 10-11, the results show that the majority of participants were able to distinguish the difference between mass and weight. On the other hand, the pre-test results of item 12 show that the participants were distracted by the visual observation and disregarded the fact that both bodies are at equilibrium. However, the game interventions were able to improve the post-test results by 5% and 12% for with tutor condition and without tutor condition, respectively. As for force and torque (item 13 - 14), the game interventions marginally improved the participants understanding in both conditions.

Table 5.3: Misconceptions in physics.

Principle	Id	Item	Answer	with tutor Pre-test	with tutor Post-test	without tutor Pre-test	without tutor Post-test
Newton's first law	1	cause of a book in stationary	(a) gravity	14.29%	19.05%	17.65%	23.53%
			(b) net force is zero	71.43%	61.90%	70.59%	70.59%
			(c) the table keeps the book stable	14.29%	19.05%	11.76%	5.88%
	2	the total force of a car with constant speed	(a) has the same direction as the car	42.86%	38.10%	23.53%	17.65%
			(b) depends on the car speed	19.05%	4.76%	11.76%	17.65%
			(c) is zero	38.10%	57.14%	41.18%	64.71%
			(d) is equal to the car weight	0%	0%	11.76%	0%
			(e) depends on the car mass	0%	0%	11.76%	0%
	3	the net force of a car with constant speed and direction to the right	(a) **is zero**	38.10%	47.62%	52.94%	47.06%
			(b) has the same direction with the car	61.90%	52.38%	41.18%	47.06%
			(c) has the opposite direction to the car	0%	0%	5.88%	5.88%

Continued on next page

Table 5.3 – continued from previous page

Principle	Id	Item	Answer	with tutor Pre-test	with tutor Post-test	without tutor Pre-test	without tutor Post-test
Newton's second law	4	A cart pushed with constant horizontal force and filled slowly with rain water	(a) has constant acceleration	23.81%	19.05%	23.53%	17.65%
			(b) continuously de-accelerates	71.43%	80.95%	58.82%	82.35%
			(c) has constant speed	4.76%	0%	17.65%	0%
	5	Two identical boxes on the earth and the moon achieved the same acceleration	**(a) The applied force of equal magnitude**	14.29%	19.05%	23.53%	29.41%
			(b) The applied force bigger on the earth	71.43%	66.67%	52.94%	58.82%
			(c) The applied force bigger on the moon	14.29%	14.29%	23.53%	11.76%
Newton's third law	6	The reaction of acting force of a pot that lying on the table with downwards direction	(a) the force from the earth to the pot	19.05%	4.76%	17.65%	0%
			(b) the force from the table to the pot	38.10%	52.38%	47.06%	64.71%
			(c) the weight of the pot to the earth	42.86%	42.86%	35.29%	35.29%
	7	The reaction of the weight of a box hanged to the roof with a rope	(a) The force from the box to the rope	19.05%	23.81%	11.76%	11.76%

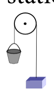
Continued on next page

Table 5.3 – continued from previous page

Principle	Id	Item	Answer	with tutor Pre-test	with tutor Post-test	without tutor Pre-test	without tutor Post-test
			(b) The force from the roof to the rope	0%	0%	0%	0%
			(c) The force from the rope to the box	19.05%	33.33%	23.53%	35.29%
			(d) The force from the box to the earth	61.90%	42.86%	64.71%	52.94%
	8	Pulling the laces of your shoes when you are on a balance	(a) decreases the value shown by the indicator	23.81%	14.29%	17.65%	11.76%
			(b) increases the value shown by the indicator	38.10%	47.62%	29.41%	35.29%
			(c) remain the same	38.10%	38.10%	52.94%	52.94%
The concept of force	9	a golf ball is moving in the air after being knocked; it has the following acting forces	(a) the gravity only	0%	0%	5.88%	5.88%
			(b) the gravity and the knock	19.05%	19.05%	5.88%	11.76%
			(c) the gravity and the air resistance	38.10%	38.10%	47.06%	23.53%
			(d) the knock and the air resistance	4.76%	0%	0%	11.76%

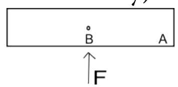
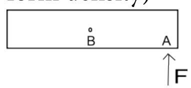
Continued on next page

Table 5.3 – continued from previous page

Principle	Id	Item	Answer	with tutor		without tutor	
				Pre-test	Post-test	Pre-test	Post-test
			(e) the gravity, the knock, and the air resistance	38.10%	42.86%	41.18%	47.06%
Mass vs. weight	10	A stone is weighted on the earth and the moon	**(a) the weight on the earth is bigger**	80.95%	85.71%	94.12%	88.24%
			(b) the weight on the moon is bigger	9.52%	4.76%	5.88%	5.88%
			(c) the weight on both places are the same	9.52%	9.52%	0%	5.88%
	11	A stone was weighted on the earth and another stone was weighted on the moon; both showed the same weights	(a) the stone on the earth has bigger mass	19.05%	33.33%	5.88%	17.65%
			(b) the stone on the moon has bigger mass	66.67%	52.38%	64.71%	70.59%
			(c) both have the same masses	14.29%	14.29%	29.41%	11.76%
	12	The bucket and box are in stationary position 	(a) the box has bigger weight	80.95%	71.43%	64.71%	47.06%
			(b) the bucket has bigger weight	4.76%	9.52%	5.88%	11.76%

Continued on next page

Table 5.3 – continued from previous page

Principle	Id	Item	Answer	with tutor		without tutor	
				Pre-test	Post-test	Pre-test	Post-test
			(c) both have the same weight	14.29%	19.05%	29.41%	41.18%
	13	Applying force F to B (uniform density) 	**(a) it will shift**	47.62%	52.38%	70.59%	70.59%
			(b) it will rotate	33.33%	23.81%	17.65%	11.76%
			(c) both shift and rotate	0%	14.29%	0%	0%
			(d) nothing will happen	19.05%	9.52%	11.76%	17.65%
	14	Applying force F to B (uniform density) 	(a) it will shift	14.29%	9.52%	5.88%	5.88%
			(b) it will rotate	38.10%	52.38%	41.18%	58.82%
			(c) both shift and rotate	42.86%	38.10%	47.06%	35.29%
			(d) nothing will happen	4.76%	0%	5.88%	0%

* *bold font indicates the correct answer*

C. Subjective learning

Beside quantitatively measuring the learning outcomes, we also measured subjective learning experienced by the participants (qualitative). This represents the feeling of participants about their learning experience. The means and standard deviations are shown in Figure 5.5, with tutor condition: 5.88(±1.12) is slightly higher compared to without tutor condition: 5.35(±0.99). One way ANOVA with the subjective learning as dependent variable and conditions as factor, shows no significant effect with $F(1,37) = 2.27$, $p = 0.14$.

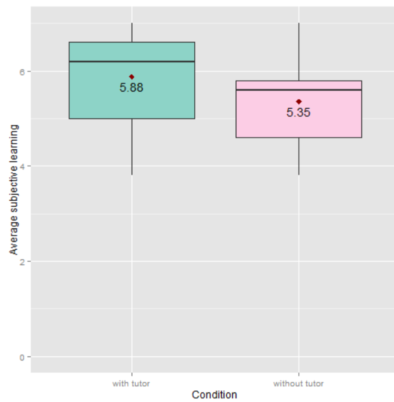
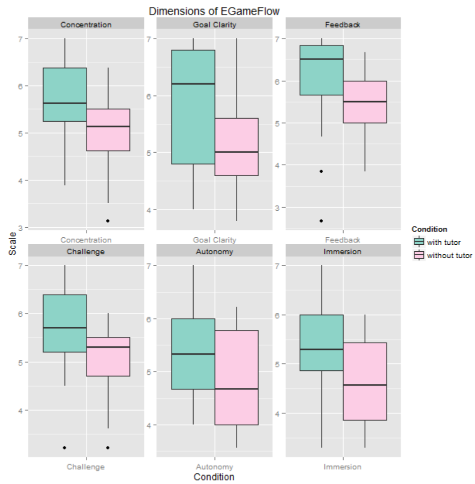


Figure 5.5: Subjective learning for both conditions.

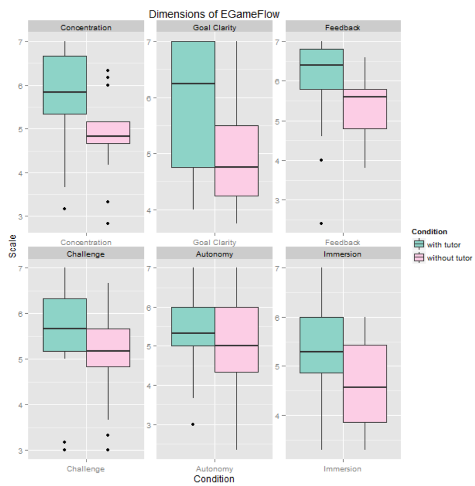
5.4.2 Flow

We measured flow using EGameFlow questionnaire minus the indicator for social interaction. In this experiment, the score of each indicator was the average of scores in all question items corresponding to it. For instance, the score for immersion is the average of scores in 7 items, i.e. I1 to I7 (Table A.2 of Appendix A). In addition, both the adapted EGameFlow questionnaire (50 items, without social interaction) and the EGameFlow questionnaire (36 items) were used to investigate the difference between both versions in measuring flow. Figure 5.6 shows the scores of the indicators in both conditions using both questionnaires, which exhibit similar tendencies, i.e. with tutor condition has higher scores in all indicators compared to without tutor condition. One way ANOVA for each indicator as dependent variable and conditions as factor using EGameflow questionnaire (36 items), shows significant effects in concentration with $F(1,37) = 6.23$, $p < 0.05$; goal clarity with $F(1,37) = 7.07$, $p < 0.05$; and immersion with $F(1,37) = 6.29$, $p < 0.05$; marginal effects on feedback with $F(1,37) = 3.37$, $p < 0.1$; and challenge with $F(1,37) = 3.08$, $p < 0.1$. There is no significant effect on autonomy with $F(1,37) = 0.62$, $p > 0.1$. This tendency also holds for adapted

EGameFlow version (50 items), except in autonomy which has marginal effect with $F(1,37) = 3.13$, $p < 0.1$.



(a) Average scores for each indicator with 50 items

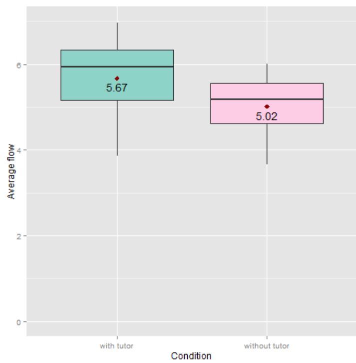


(b) Average scores for each indicator with 36 items

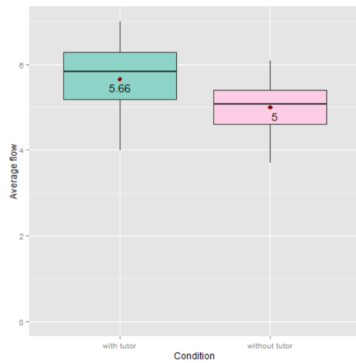
Figure 5.6: Average scores for each indicator in both versions of EGameFlow questionnaire.

The GameFlow model used all indicators above to measure flow in games (Sweetser and Wyeth, 2005). This is done by averaging all indicators, including concentration, goal clarity, feedback, challenge, autonomy, and immersion. Figure 5.7 shows the score for flow in both conditions using two versions of EGameFlow questionnaire. The scores for average flow are 5.67(±0.83) for with tutor condition and 5.02(±0.84) for without tutor using the adapted EGameFlow (50 items); 5.66(±0.84) for with tutor condition and 5.00(±0.72) for without tutor using the EGameFlow (36 items).

One way ANOVA for flow as dependent variable and conditions as factor using EGameFlow questionnaire (36 items), shows significant effects with $F(1,37) = 6.55$, $p < 0.05$. This also holds for adapted version with 50 items with $F(1,37) = 6.71$, $p < 0.05$.



(a) Average flow with 50 items



(b) Average flow with 36 items

Figure 5.7: Average flow in two versions of EGameFlow questionnaire.

5.5 Discussion and conclusions

Misconception

The results of the pre-test show that the majority of the participants have an acute misconception in Newtonian mechanics, in particular associating the motion of a body to a constantly acting force on the body with the direction of the motion. This tendency is common among students globally and has been the subject of many studies (Stylos et al., 2010). We were expecting the gaming conditions -which were designed to address misconceptions by providing a game to simulate phenomena at a macroscopic level, hints and feedback (including a tutor) to explain the phenomenas at conceptual level- to be able to reduce this misconception among the participants. The treatment using the game with the tutor was able to correct the perception of some participants although half of them still preserved their misconception.

Though most of the participants have a rather good understanding of the relation between weight and mass, there was confusion among them in the meaning of the weight of a mass. This led to incorrect perception that the heavier weight of a body due to the earth gravity requires larger force to accelerate the body compared to if the body lies on the moon. Several problems in the games addressed this issue, but the intervention only improved the number of correct responses in a small percentage; 70-80% of participants retained their incorrect understanding. The majority of the participants also misunderstood asymmetric equilibrium in which the sum of acting force on the system is zero (item 12). They were distracted with the visual observation in form of static image and posited that the object closer to the earth is heavier.

From the pre-test and the post-test, we could notice that small number of participants know Newton's third law and the intervention in both conditions were able to correctly introduce it. However, the participants failed to apply it directly to a real situation even after the intervention. The simulation in the game depicted the acting force and its corresponding reacting force. In addition, the game provided hints and feedback . However, it turned out to be ineffective. Introducing Newton's third law requires natural context of interaction between two bodies which probably using real world experiments may help the students in understanding the concept. Several studies also have vindicated similar difficulties in teaching Newton's third law to the students (Terry and Jones, 1986; Brown, 1989; Savinainen et al., 2005; Stylos et al., 2010).

Flow

We hypothesized that between two conditions, the one with the tutor would obstruct the game continuity and results in lower flow. However, our finding controverted the hypothesis. The one with the tutor received significantly higher level of flow compared to the one without the tutor. The participants saw the tutor as helpful in the process of problem solving and it was perceived as an integral part of the game. We concluded that infusing scaffolding in the form of a tutor into gaming does not need to obstruct flow. One reason could be drawn from (Paras and Bizzocchi, 2005). They argued that games can act as effective learning environments by integrating reflection into the process of play, producing an endogenous learning experience that intrinsically motivating. This emphasizes the importance of reflection to prevent the learner from wandering around aimlessly and to fully realize learning. However, the reflection period may disrupt flow. To avoid this, reflection must appear to the learner as one of the many in-game goals that drive the game-play. This was most likely the case for the tutoring tool in our implementation as it was designed as part of the mission in the game. Furthermore, games are fun because we learn new things (Koster, 2013) and most likely that reflection in games that enables learning new things are also fun. This work is useful for educators in implementing a tutoring system in games that effectively improves flow.

In this experiment, we found concentration, goal clarity, immersion, feedback, and challenge to be important indicators that distinguish both conditions. The tutoring tool, which was implemented as part of feedback mechanism, contributed to the significant difference of feedback between to games. In addition, with different questionnaire, Kiili (2006) found that there is a connection between goal clarity and feedback to flow. Therefore, our finding further supports this connection. We could not find significant effect of autonomy between our games. Similar GUI and game interaction style in both conditions might contribute to insignificant difference in autonomy.

Learning and Flow

In the light of our findings, we concluded that two gaming conditions have significantly different level of flow but they have no significant effect on the learning outcomes. The results of our study are in contrast to the previous research that indicated a loose positive connection between flow and learning in games (Kiili, 2006). The main reason is that the previous research did not measure the learning outcome in quantitative manner but rather in qualitative manner via interviews. In this regards, we found similar tendency in the subjective learning; with tutor condition -that has higher level of flow- shows slightly better feeling of learning, although the difference is insignificant. Hence, it may suggest a loose relationship between flow and the feeling of learning. Other possible reasons of insignificant effect of flow on the learning outcomes might be that the difference in flow between two gaming conditions were not high enough to induce different level of learning, or we did not completely capture the whole dimensions of flow and learning.

Limitations and notes on the experimental setup

There are four possible threats to the validity of the results. Firstly, the small number of knowledge test items; fourteen items for conceptual knowledge and eight items for procedural knowledge may be too few to accurately depict the effects of games in learning. However, too many test items would also compromise the learning outcomes due to long concentration needed to perform the test. Secondly, all the materials used in the games were in English, whereas all participants were non native English speakers. This might affect the test results due to both the unfamiliarity with physics terminologies in English and the differences in perceiving the terminologies. This especially holds for dealing with misconceptions. Moreover, the abusive use of everyday language may alter the interpretation of the terminologies and this should be considered as potential source of misconceptions (Stylos et al., 2010). Thirdly, since we performed a post-hoc evaluation of flow, it is unclear whether flow was directly associated with learning, or whether flow levels were simply not high enough to see a real effect on learning, which in turns it needs to be stimulated more.

Fourthly, while games are played freely, serious games in class are played on a compulsory basis, which may change the user attitude and perception while playing.

5.6 Future works

This works provided a baseline for researchers working in the area of serious games for education, in particular how a tutoring system as a feature/service could be useful in a physics-based simulation game. We measured both flows and the learning outcomes (subjective and objective measures) in games based learning and concluded that flow does not necessarily correlate to learning as previously stated. Conventional teaching that views learning as a painful process with a well-known adage "no pain no gain" seems to support this, in particular for deep learning (Graesser et al., 2009). However, our finding also did not confirm that lower flow in games promotes better learning. Therefore, future research could investigate this further using different types of games in various topics.

Our work only focused on one feature in games, i.e. scaffolding in the form of tutoring tool. Other game artefacts, such as narrative and fantasy, and playing-styles can be further investigated in relation to flow and learning. This will benefit serious games designers and developers in tuning their games and game features for better learning. This will also support the development of ready-to-use services for serious games development.

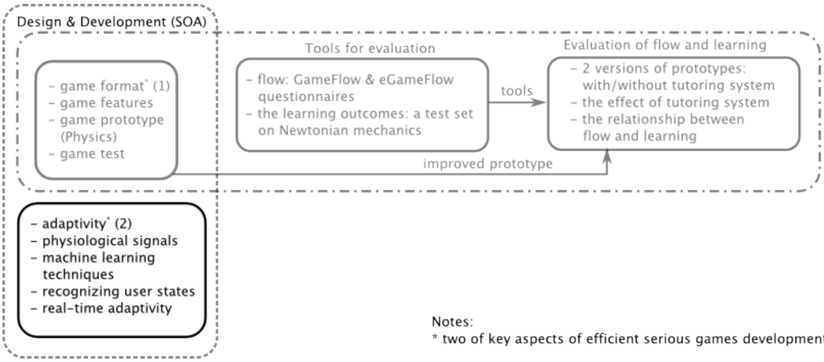
Results may suggest that flow in serious games is less relevant than in entertainment games. Probably, different, or not all the dimensions of flow could be considered for characterizing serious games. Future research could investigate further on the dimensions of serious games and the importance of the dimensions for characterizing serious games. Although flow may not affect real learning, serious games should be fun for other reasons than learning efficacy, e.g. to motivate bad/under-performed students, to assist unaware students to learn a new topic, and also to improve the feeling of self-efficacy and self-competence. This may make students more eager to continue learning (Van der Spek, 2012). This requires adaptivity to control challenges (e.g., game speed, score, music) in games. This issue is further investigated in the subsequent part (i.e., Part III).

Part III

Adaptivity as a Service in Games

Chapter 6

Physiological signals and affective states



"Feelings aroused by the touch of someone's hand, the sound of music, the smell of a flower, a beautiful sunset, a work of art, love, laughter, hope and faith - all work on both the unconscious and the conscious aspects of the self, and they have physiological consequences as well." - Bernie Siegel

Abstract. Computational devices have become more intelligent in terms of their abilities to perceive the environment and to respond accordingly. This has been enabled by the sensor technologies with the aid of machine learning algorithms. As more practical physiological sensors are becoming available on the market, they are not only useful in medical researches, but they also offer human-computer-interaction (HCI) and adaptation researchers new approaches in interaction. Physiological activities/signals may represent emotions which we may exploit to recognize human affective states. This is highly desirable for providing empathetic responses to human or infusing social competencies into devices. In gaming context, physiological signals provide users with new ways of interaction and/or content adaptation to better suit their preferences. Therefore, this chapter presents several physiological signals that may be useful for researches in both HCI and adaptation mechanism.

6.1 Introduction

In the previous part, we have designed, developed, and evaluated a game and game features (services). We found that the implementation of a tutoring system improves flow, albeit no effect on real learning. Moreover, we found no evidence on flow improves learning. However, it does not necessarily mean that flow is not important. In general, any serious game should provide fun for other reasons than learning efficacy, e.g. to improve the feeling of self-efficacy and self-competence which may make students more eager to continue learning (Van der Spek, 2012). This still requires adaptivity mechanisms to control level of challenges in games. Moreover, adaptivity has been shown to lead to more efficient serious games in terms of learning over time (Van Oostendorp et al., 2014). However, this work was based on an in-game scoring method which is not always suitable for all serious games and only an indication for coping level. On the other hand, measuring physiological signals, in particular brainwaves, is a more direct indication of experienced challenge, which makes a flow-related neuro-physiological characterization of a player very important. Consequently, in this part, we directed our focus on the adaptivity to support efficient games development in general using physiological signals. This starts with reviews on physiological signals and human affective states.

Computers today have become more intelligent. Not only they come with different shapes and sizes, e.g. tablets, smart-phones, smart-watches, and smart-homes; they are also increasing in terms of capabilities by being able to perceive their environment using sensor technologies and to respond accordingly with the aid of machine learning algorithms. This is not only useful for improving human-computer interaction (HCI) in a more natural way, such as swiping an e-book page in a touchscreen to move between pages, but also for adapting the environment according to the presence of human, e.g. adjusting the room temperature.

In essence, both HCI and adaptation mechanisms rely on the use and perception of human behavioral signals such as speech, motions, gestures that communicate intentions to a computer. For instance, person A shouts and his face turns to red, exhibiting anger in front of other person, e.g. B. In communication perspective, the sender, A, communicates his state in verbal and behavioral forms and the receiver, B, senses the state of A by decoding the verbal and behavioral cues that he perceives based on his own experiences (Figure 6.1). This type of communication is sequential, where feedback may be provided to the sender but non-simultaneously (Wood, 2011). This is commonly seen in human computer interaction such that systems sense the users' requests/states using sensor technologies and the systems respond to the users accordingly. In this case, both the users and the systems may learn through experiences of interacting between each other, and may offer multiple sensing modalities. For instance, in a gaming context, computers may sense users via, for instance, the microphone, the camera, or haptic devices, and in return, they may give feedback in the form of visual responses, audio feedback, or haptic feedback (Andrews et al., 2006; Rosenberg and Riegel, 2002; Piekarski and Thomas, 2002).

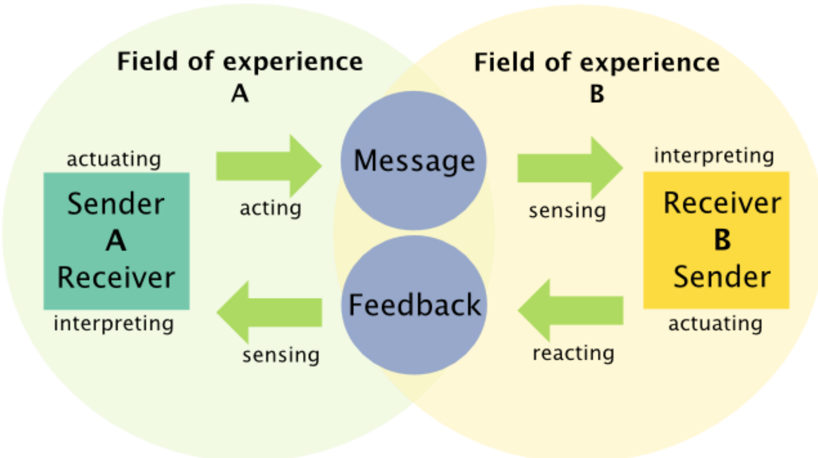


Figure 6.1: Interactive model of communication

However, to interact naturally with humans, computers must recognize human affective states and express social competencies. This gave birth to a new area in computer science, i.e. affective computing, which studies and develops systems and devices that can recognize, interpret, process, and express human affects (Tao and Tan, 2005). This requires multidisciplinary studies spanning between computer science, psychology, and cognitive science.

There are two stages of study in affective computing: a) detecting and recognizing emotional information (sensing and interpreting), b) embodying emotional intelligence to machines (sensing, interpreting, reacting/empathetically responding). Detecting emotional information begins with sensor technologies that capture data about the user's physical state or behavior. This is analogous to the way humans collect cues to perceive emotions in others. For instance, a video camera captures facial expressions, body posture and gestures, a microphone captures speech/utterance; and physiological sensors detect emotional cues by directly measuring physiological data, such as skin temperature and galvanic resistance (Garay et al., 2006). Subsequently, recognizing emotional information requires techniques in machine learning to extract meaningful patterns from the data, such as a stammered utterance that may express fear, and bulging eyes, bulging veins, a red-face, and a higher-than-normal pitched voice may express anger. The subsequent stage, i.e. embodying emotional intelligence to machines, requires the design of computational devices to exhibit either innate emotional capabilities or are capable of convincingly simulating emotions. This aims at a more empathetic interaction between computers and humans, such as conversational agents with cultural and emotional intelligence (Heise, 2004).

Our focus, however, is on the first stage, i.e. to detect and recognize the users' affective states for adaptation mechanisms, by using physiological sensors, particularly in a gaming context that engenders high stimulation. Therefore, the subsequent sections discuss several physiological data that may be useful for achieving our goal, in particular brain activity, heart rate, and skin conductivity.

6.2 Brain activity

The first attempt in recording electrical activity of the human brain was performed by Hans Berger in 1924 using his invented device called electroencephalogram (EEG), the results of which were published in 1929 (Berger, 1929; Haas, 2003). This marked a momentous advancement in clinical neurology. The EEG measures the electrical charges of the neurons in the human brain by use of electrodes placed on the scalp. The ions exchange either between neurons, or neuron and metal electrodes, can be measured by a voltmeter. Hence, recording these voltages over time gives us the EEG. In Berger's experiment, two large sheets of tinfoil in the forehead and in the back of the head were used to serve as electrodes, whereas now, EEG with arrays of electrodes is a clinical routine in brain research.

6.2.1 Instruments and objectives

EEG recordings become more prevalent with the addition of other emerging acquisition techniques relying on electrical, magnetic, and haemodynamics (blood circulation) activities of the brain. Magnetoencephalography (MEG) studies the magnetic signals associated with the electric currents of the brain (Cohen, 1968; Hämäläinen et al., 1993). This magnetic fields are measured by a superconducting quantum interference device (SQUID), a sensitive detector of magnetic flux invented circa 1960's (Zimmerman et al., 2003). Brain anatomical structures can be investigated by computer aided tomography (CT) scans and by magnetic resonance imaging (MRI) which provide high-quality images of the brain tissues but not the metabolic information of the brain. Brain activity, in the form of blood circulation and oxygenation, can be measured further with very good spatial accuracy using nuclear imaging techniques, e.g. positron emission tomography (PET) (Jasczak, 1988), functional magnetic resonance imaging (fMRI) (Belliveau et al., 1991), or using opto-electromagnetic instruments, e.g. near-infrared spectroscopy (NIRS) (Treado et al., 1992).

PET has a temporal resolution on the order of seconds or around 0.1 second at best, whereas fMRI data has longer intervals (within 100-ms intervals) with the limitation of 1-s for the blood flow in the brain (Tanzer et al., 2006). On the other hand, electro-magnetic approaches provide better temporal resolution. For instance, MEG has approximately 1-ms temporal resolution. In addition, a very important advantage of the latter techniques is that they are non-invasive, whereas, in nuclear imaging techniques, researchers need to consider the maximum radiation dosage to safeguard the subject under examination. Gürkök and Nijholt (2012) listed and compared four non-invasive methods used in measuring brain activity, in particular the pros and cons of each method (Table 6.1).

Table 6.1: Comparison of methods for measuring brain activity (Gürkök and Nijholt, 2012)

Properties	EEG	MEG	fMRI	NIRS
Measured activities	electrical	magnetic	haemo-dynamic	haemo-dynamic
Temporal resolution	high	high	low	low
Spatial resolution	low	low	high	low
Portability	high	low	low	high
Cost	low	high	high	low

Among those methods in the HCI area, EEG and NIRS are practical and feasible to be used for developing brain computer interface (BCI), i.e. a communication system in which intentions of an individual are captured in the form of brain activity and it produces supporting actions according to the intention/psychological state of the brain activity (Wolpaw et al., 2002; Gürkök and Nijholt, 2012). This is because EEG and NIRS are portable and inexpensive to deploy. In gaming context, this reason is also valid, in particular for EEG, since EEG is the most common and accessible brain activity measuring tool in the game market (e.g. Neurosky[1], Emotiv[2],IntendiX[3], MindGames[4], Mattel[5]). Our focus, however, is not on the use of EEG for BCI, but more on to the use of the EEG as a tool for recognizing human affective states in gaming (affective computing).

Thus, we may categorize the application domains of EEG for BCI and affective compute areas into two types: passive EEG and active EEG. In active EEG, the user interacts with the EEG application to directly control it, whereas in passive EEG, the user is only being monitored to adapt the task or the environment according to the user's condition (Gürkök and Nijholt, 2012).

[1] http://www.neurosky.com/
[2] http://www.emotiv.com/
[3] http://www.intendix.com/
[4] http://mindgames.is/
[5] http://mindflexgames.com/

6.2.2 EEG rhythmic activity

During a normal condition in which no intervention is given, the brain maintains a neural oscillation (rhythmic activity). This rhythmic activity is divided into frequency bands and each frequency band may occur with higher/lower amplitude in different areas of the brain. In case of an internal/external event, suppression or enhancement of the rhythmic activity may happen which is referred as event related desynchronisation (ERD) and event related synchronisation (ERS) (Pfurtscheller et al., 2006). Therefore, by observing the signal amplitude in certain frequencies measured at specific parts of the brain, we can infer the underlying brain activity. For instance, alpha (or alpha wave) is the frequency range between 8 Hz and 12 Hz found by Hans Berger in posterior regions of the head, on both sides of the brain hemisphere. The alpha wave emerges with closing of the eyes and with relaxation, whereas it attenuates during mental exertion (Deuschl and Eisen, 1999). Another example, beta waves frequencies from about 13 Hz to about 30 Hz are characteristic of an alert state of consciousness, whereas beta activity at even higher frequencies has been observed in different types of mental activities (Levin, 2000).

Among the basic waveforms are the alpha, beta, theta, and delta rhythms (Blume et al., 2010; Fisch and Spehlmann, 1999; Niedermeyer and da Silva, 2004). As mentioned before, alpha waves occur at a frequency of 8 to 12 cycles per second in a regular rhythm and they are present if we are awake but with our eyes closed. Beta waves occur at a frequency of 13 to 30 cycles per second and they are usually associated with anxiety, depression, or the use of sedatives. Theta waves occur at a frequency of 4 to 7 cycles per second and they are seen in sleep. Delta waves occur at the lowest frequency level between 0.5 and 3.5 cycles per second and they generally occur during deep sleep. Delta waves often have the largest amplitude among all brain waves. Other signals include gamma and mu which correspond to higher mental/cognitive task and sensorimotor task, respectively (Pulvermüller et al., 1997; Pfurtscheller et al., 2006). Therefore, monitoring the activities of these frequency bands may provide us an alternative to recognize the player's state and to adapt the task according to the user's condition.

6.3 Other physiological activities

Beside brain activity, physiological cues that may be useful to infer human affective states include: skeletal muscles activity, respiratory volume (RV), blood volume pulse (BVP) and heart rate (HR), skin temperature, and skin conductance/galvanic skin response (GSR) (Gouizi et al., 2011). Skeletal muscles activity, including facial muscles, can be measured by using electromyogram (EMG), which records the electrical potential generated by muscle cells. Skeletal muscles, specifically facial muscles, are sensitive to emotional reactions such as fear (Dimberg, 1990). RV and BVP correspond to changes in air and blood volume within an organ or whole body and they can be measured using plethysmograph.

HR is the number of heartbeats per unit of time which can be measured using either electrocardiogram (ECG) or photoplethysmograph (PPG). ECG, invented by Willem Einthoven (Einthoven, 1902), is widely used in clinical setting, such as for studying the heart rhythm, diagnosing arrhythmias, studying the metabolism of the heart, and assessing cardiovascular risk due to occupational hazards (Kligfield et al., 2007). ECG involves several electrodes to be placed on the surface of the thorax or on the limbs. This makes ECG less practical outside of clinical settings. On the other hand, PPG, a form of optical plethysmograph, is efficient since it uses a light-emitting diode (LED) to illuminate the skin (placed on finger/toe) and then measuring the amount of reflected/transmitted light to a photodiode (Nijboer et al., 1981).

The normal resting adult human heart rate ranges from 60-80 beats per minute (bpm) (Palatini, 1999). Heart rate abnormalities may be caused by pathological conditions, physical conditions, and emotional conditions. For instance, Tachycardia, a fast heart rate, may occur due to fever, physical exercise, or anxiety. Bradycardia, a slow heart rate, may occur due to regular exercise (Palatini, 1999). This means, assuming a healthy subject, an increasing heart rate can be introduced by stimuli that promote emotional anxiety, which can be useful for detecting and recognizing anxiety.

Skin conductance (GSR) measures the electrical conductance of the skin which varies due to rapid fluctuations in eccrine sweat gland activity which is controlled by the sympathetic nervous system (Boucsein, 2012). Therefore, skin conductance is used as an indication of psychological/emotional or physiological arousal (Lanzetta et al., 1976; Cuthbert et al., 2000). Arousing the sympathetic branch of the autonomic nervous system increases the sweat gland activity and consequently, increases skin conductance. In this sense, skin conductance can be used as a measure of emotional and sympathetic responses (Carlson, 2012). Skin conductance can be efficiently measured using GSR electrodes placed on the fingers or palms.

Among those physiological activities, HR and GSR are relatively practical to measure, in particular for gaming. Moreover, they represent functions of experience that indicate anxiety and stress level (Fenz and Epstein, 1967).

6.4 The use of physiological signals in detecting and recognizing human affective states

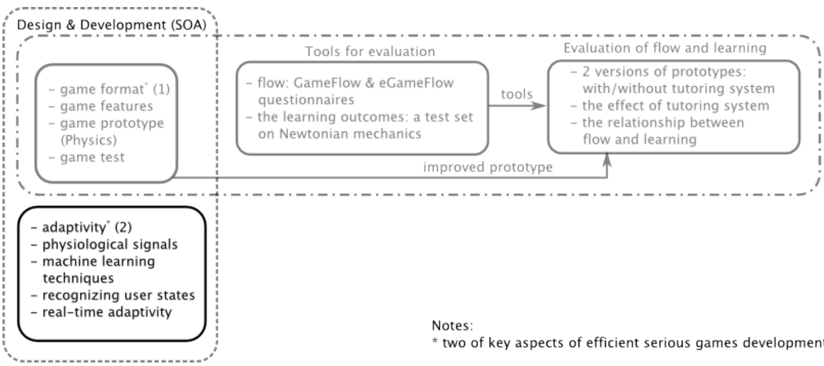
Many papers reported human physiological characteristics as one of the approaches to recognize emotions and human affective states. Mandryk and Atkins (2007) developed fuzzy logic models that transformed physiological signals into arousal and valence to quantify emotions. Arroyo et al. (2009) used physiological sensors, such as facial, seat, and wrist expressions, to predict four different classes of feeling: confident, frustrated, excited, and interested, while learning with a tutoring system. They achieved 60% accuracy and provided evidence that modifying the context of the tutoring system using the captured feeling may optimize students' emotion reports and in turn improve math attitudes. Likewise, Brawner and Goldberg (2012) used electrocardiogram (ECG) and galvanic skin response (GSR) to monitor the physiological state of learners in computer-based training with and without intervention. The results show the appropriateness of instructional intervention with respect to excitement and provide insight in the development of real-time assessment.

In HCI and adaptation mechanisms, physiological measures have been used extensively to improve the user experience. For instance, Wilson and Sasse (2000a,b) used GSR and ECG to examine subject response to the quality of audio and video in video conferencing software. Scheirer et al. (2002) used pre-programmed mouse delays to intentionally frustrate a computer user which then applied Hidden Markov Models (HMMs) to GSR and blood volume pulse (BVP) data to detect states of frustration. Mandryk and Inkpen (2004) showed that GSR and EMG of the jaw were higher when playing a computer game against a friend over playing against a computer. They also found strong correlations between GSR and fun, and EMG and challenge. Hence, physiological measures provide a rich, continuous, and objective source of information about the user experience. Subsequently, Mandryk et al. (2006a) developed a fuzzy physiological approach for modeling emotion during interaction with play technologies using GSR, HR, and EMG. The result showed a great potential of using physiological metrics to model emotional experiences for interactive play technologies. Pun et al. (2006) set up researches in physiological signals, in particular EEG, for multimodal interaction between humans and computers. In gaming context, Gürkök and Nijholt (2012) studied principles of active BCI for multimodal interfaces in gaming.

Although many papers reported the potential of physiological signals to infer a user's affective state (user experience), the applications in games are still limited, in particular for difficulty adaptation. Moreover, previous studies did not assess the possibility of using physiological signals for real-time adaptation. Among those physiological signals, EEG has not been well exploited though recently it has become more practical as shown by various products available in the market. Therefore, in our work, we aimed at real-time difficulty adaptation using physiological signals, in particular EEG, HR, and GSR, as described in Chapter 7.

Chapter 7

Adaptivity in games using physiological signals



"If little else, the brain is an educational toy." - Tom Robbins, *Even Cowgirls Get the Blues*

Abstract. Computer games are very popular form of entertainment with a variety of genres and consumer groups. Due to their ability to capture the player's attention for a long period of time, researchers in the area of pedagogy have been investigating the use of computer games for learning. Despite the popularity of computer games, researchers still suffer from a lack of effective evaluation methodologies to verify player engagement with games. Using an empirical approach, this chapter is aimed at investigating the emerging physiologically based human-computer interaction (HCI), to identify Csikszentmihalyi's three emotional states of players in gaming: boredom, flow, and frustration, for difficulty adaptation. To this end, we collected physiological data of players during gameplay using Elemaya[1], a biofeedback instrument with 4-electrodes of electroencephalogram (EEG), 1 channel of skin conductance (GSR) and 1 channel of photoplethysmogram for heart rate (HR). Subsequently, we used techniques in machine learning to estimate the player's states from the physiological data[2].

[1] http://www.elemaya.it/
[2] this chapter is based on (Berta et al., 2013; Plotnikov et al., 2012)

7.1 Difficulty adaptation

Computer games are known for their ability to capture the player's attention for a long period of time. This inherent game property has attracted many psychologists to study the enjoyment factors of games, which in turn engendered several theoretical frameworks on enjoyment in games.

The foremost theory that describes the optimal experience in games is Csikszentmihalyi's theory of flow explained in Chapter 1. This theory states that the level of challenge in a game should match with the player skills. Consequently, an enjoyable game by definition is a game that adapts its level of challenge to the skill of its audience. Likewise, the GameFlow model defines the enjoyment criteria to include concentration, challenge, player skills, autonomy, clear goals, feedback, immersion, and social interactions (Sweetser and Wyeth, 2005). These criteria were used in the flow framework as the flow antecedents, i.e. factors that contribute to flow (Kiili, 2005a). Another prominent theory of engaging game play is Malone's principles: challenge, curiosity, and fantasy (Malone, 1981). Likewise, Lazzaro introduced four elements of entertainment in games: a) hard fun corresponds to challenge in Malone's principles; b) easy fun corresponds to curiosity in Malone's principles; c) altered states correlates to internal emotions and fantasy in Malone's principles; d) socialization represents social interactions (Lazzaro, 2004). This shows similarities of several enjoyment theories in games.

Based on the factors of flow, researchers have been working on quantifying those factors into numerical models. (Yannakakis and Hallam, 2007) estimated general criteria of interestingness in the Pac-Man game, a predator/prey game genre where the prey is controlled by the player and the predators are controlled by the computer. Three metrics were invented by focusing on the challenge, i.e. the behavior of the opponents: a) level of challenge is numerically represented by the average number of steps taken by the opponents to kill the player in a long run, b) behavior diversity of the predators is manifested in the variance of time taken to kill the player, and c) spatial diversity of the predators is represented by the randomness of predators' movement using average entropy values. Despite their practical application, there are two assumptions that limit the merits of the metrics. Firstly, the metrics assumed the interaction with the opponents is the primary source of variance in enjoyment. Secondly, the metrics assumed an average level of playing skills. This means the merits of the numerical model may diminish for large number of players with various levels of skills.

Thus, designing a balance game becomes a greater challenge as the size of the potential audience grows, which is the typical case of video games. Most games presently offer only a single narrow, static experience, which might keep the typical player in flow, but may not be fun for the hardcore or the novice player alike (Figure 1.2(b) in Chapter 1). Several choice possibilities should be given to the player to adapt to different users' personal flow zones (Chen, 2007).

On the other hand, simply increasing the number of choices is costly, and an excessive number of choices risks overwhelming the user, because of the frequent interruptions. This compromises the fundamental components of flow described in the flow framework - a sense of control and concentration on the task at hand. Therefore, to avoid these counterproductive situations, designers have to embed the player choices into the core activities of the interactive experience (Chen, 2007) and/or make the game automatically adaptive (Lopes and Bidarra, 2011) by assessing player state (Chanel et al., 2011; Liu et al., 2009), which motivated our work in this chapter. One of ways is to create a reusable service for adaptivity which will work for games in general.

In this chapter, we investigated the use of the emerging physiologically based human-computer interaction (HCI) to identify Csikszentmihalyi's three emotional states of players in gaming: boredom, flow, and frustration, for difficulty adaptation. This is because most people share similar physiological traits under certain conditions. For instance, if people feel intimidated, their blood pressure, heart rate, and skin conductivity tend to rise. In Chapter 6, we have discussed the state of the art of using physiological signals in detecting and recognizing human affective states, and by exploiting this fact, we hypothesized the difficulty adaptation to be generically tractable using physiological signals. Therefore, the main idea is that, based on the physiological state of a player, if the player experiences boredom and frustration due to an unmatched game challenge to player's skill, the gaming system will automatically adapt the difficulty of the challenge by either increasing or decreasing it in the game. This supports the educator in terms of easy and efficient serious games development since the educator does not have to determine the proficiency level of every student and balance the game accordingly. To do so, we collected physiological signals of players who played a game in three difficulty levels. Subsequently, we used techniques in machine learning to build and test a classifier using data from physiological signals.

The rest of the chapter is organized as follows. Section 7.2 explains the related works to our research and Section 7.3 presents the detail of our experimental settings and data. Section 7.4 discusses the exploratory data analysis on the physiological data, whereas Section 7.5 explains the predictive analysis on the physiological data using SVM and Bayesian framework. In Section 7.6, we performed comparison of physiological signals in different tasks. Discussion and conclusions, and future works are available in Section 7.7 and 7.8, respectively.

7.2 Related work

The advancement of tools for measuring neuro-physiological activities has opened up avenues for quantitative estimation of human affective states relevant to the learning. Pellouchoud et al. (1999) observed 14 teenagers playing a video game and found similar responses elicited by increased mental load in normal adult population. Using a 115 channel EEG, they reported an increase in the amplitude of frontal midline theta (6-7 Hz), and an attenuation in both the posterior alpha wave (9-12 Hz) and in the central mu wave (10-13 Hz), with respect to an open-eyes resting condition. Likewise, Sheikholeslami et al. (2007) conducted a high-resolution (128 channels) EEG study of dynamic brain activity during video-game play with two participants playing for a total of over 60 minutes, with 3-minutes rest periods in every 10-minutes of play. Considering only alpha and theta waves, results revealed a frontal midline theta wave activity increase, and parietal alpha wave initial decrease followed by a slow increase. Derbali et al. (2011) also showed that motivational factors in a serious game seem to elicit specific physiological trends in learners, especially observable in the EEG attention ratios (i.e., the theta/low-beta wave ratio). Babiloni et al. (2007) analyzed the concurrent brain activities of small groups of people interacting in social cooperation or competition, identifying different activated areas and frequency bands, according to the different physical or reasoning activities.

In the area of *affective ludology*, a research field focused on the physiological measurement of affective responses to player-game interaction, Nacke et al. (2011) investigated the impact of level design on brainwave activity (EEG) and player experience (in the form of questionnaires). They focused on three conditions: boredom, flow, and immersion, and reported that the immersion-level design elicits more activity in the theta band, which may support a relationship between virtual spatial navigation/exploration and theta activity.

In gaming feedback, several games have been developed to use neurofeedback/biofeedback, for instance, to support the balancing of waves from the two cerebral hemispheres (Shim et al., 2007) and improve concentration (Wang et al., 2010). Coyle et al. (2011) experimented with a new BCI-based game training paradigm which enables assessment of continuous control performance.

Girouard et al. (2009) used an experimental fNIRS system to distinguish, in terms of cognitive workload, two levels of difficulty of the Pacman game. Our approach is similar, but our contribution is mainly in the following two directions: we use a portable commercial EEG and are interested to check if more levels of boredom/flow can be detected through a state of the art classifier, and we use 1 second width windows for the EEG signals to check the possibility of a real-time adaptation.

Chanel et al. (2011) investigated the use affective information in the form of EEG and peripheral nervous system, to maintain a player's involvement in games. After confirming that their three different levels correspond to distinguishable emotions (boredom, engagement, and anxiety), they trained several classifiers to automatically detect the three emotional classes in a player-independent framework. Liu et al. (2009) also presented the affect-based dynamic difficulty adjustment (DDA) to enhance gaming experience compared to DDA without affective information (i.e., only based on player's performance).

They analyzed the heartbeat, the body temperature, the EMG to infer anxiety level. Our approach is similar to Chanel et al. (2011); Liu et al. (2009) but with different window width (i.e. 1 second width window) intended for a real time adaptation, and we compared approaches and outcomes in Section 7.7.

7.3 Experiments

7.3.1 The game and conditions

We adapted an open source game, i.e. a vertical scrolling aerial combat game, and selected appropriate levels to create three emotional conditions: boredom, enjoyment, frustration. The boredom and frustration conditions were embodied in the game by scaling the game features, such as the number and the pace of enemies, and the weaponry upgrades. The flow condition, on the other hand, required tuning the game features by precursory testing the game with several subjects distinct from the experiment.

Figure 7.1 depicts the aerial combat game for boredom condition as a black background with at most two enemy planes at a time. The goal of the game was to maximize the score by killing the enemies and avoiding penalties due to being killed. Figure 7.2 shows the game for frustration condition with an overwhelming number of enemies at hand to an extent that survival was nearly impossible. Consequently, the subjects subconsciously altered the goal into survival mode by finding a good position on the screen. Nevertheless, no strategy was actually viable, so any attempt would lead to frustration.

The flow condition involved a moderate quantity of enemy planes with collectibles in the form of weaponry upgrades to increase player strengths. Every minute the number of enemies and the collectibles were adjusted to match with the progress of the player to avoid an adaptation effect, difficulty was slightly scaled every minute. The background of the game represented an aerial view of the city of Genoa taken from Google Map[3]. This might introduce confounding factors in form of visual elements but we argue that it was necessary to establish proper levels since it may stimulate fantasy and immersion.

[3]http://maps.google.com

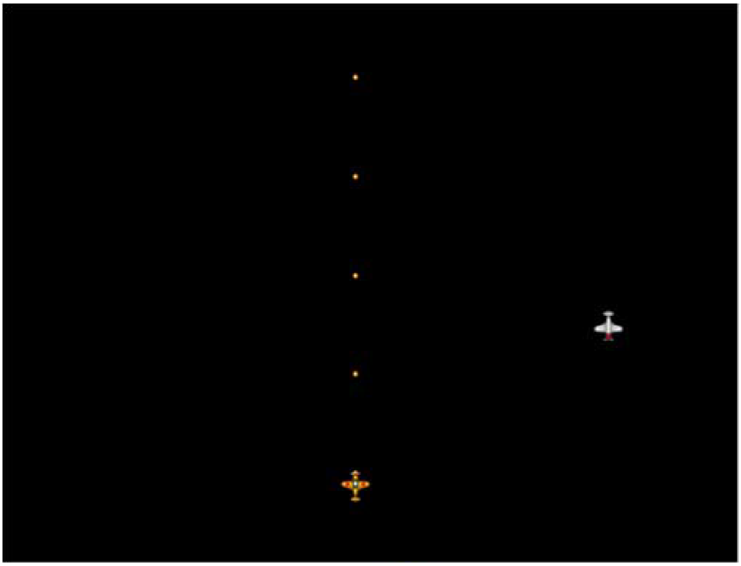


Figure 7.1: A snapshot of aerial combat game for boredom condition.

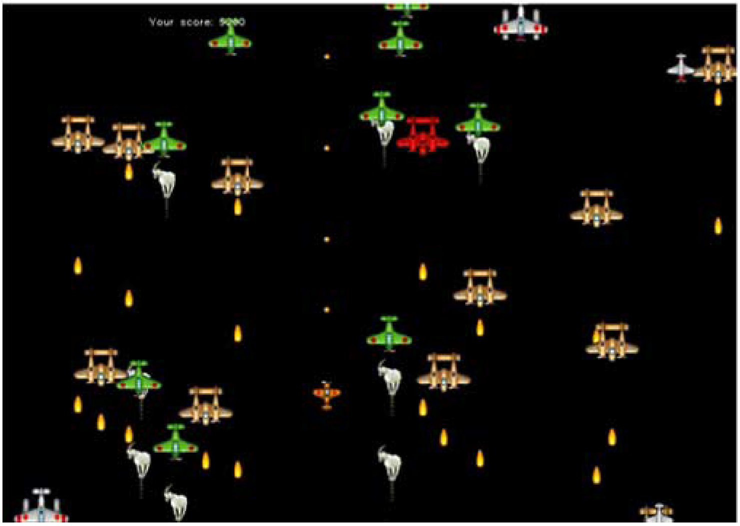


Figure 7.2: A snapshot of aerial combat game for frustration condition.

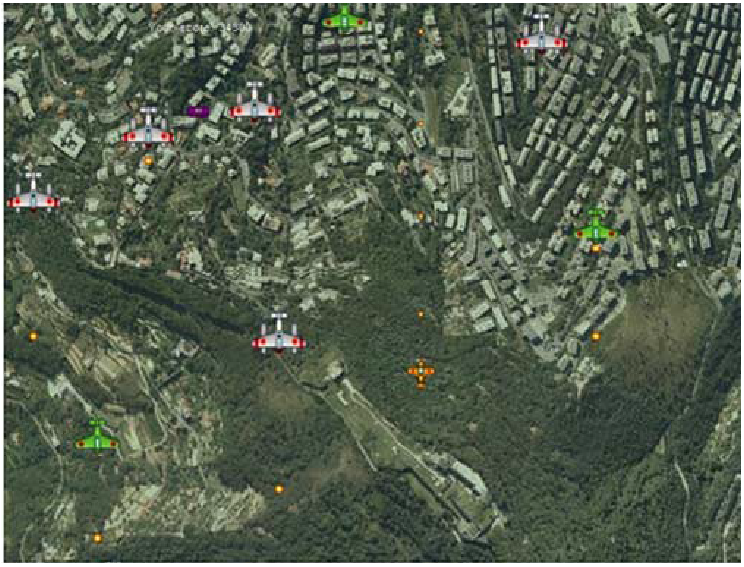


Figure 7.3: A snapshot of aerial combat game for flow condition.

7.3.2 Participants

Twenty two students, master and doctoral in the Engineering department at University of Genoa, voluntarily participated in the experiment. The mean age was 26.3 (SD = 5.5). Seventeen were male (77%) and five were female (23%). In term of game exposure, 55% claimed to be avid players, 36% were moderate players, and 9% were non game players. Seventeen participants were right handed while five of them were left handed.

Due to differences in term of playing ability, the flow condition was tuned differently based on the game experience of the players. Consequently, three sub-levels were devised in the flow condition in which each player was assigned under two criteria: a) player score on a pre-test game, a vertical scrolling obstacle avoidance car race (Figure 7.4), and b) player self assessment on his/her gaming experience.

Figure 7.4: A snapshot of pre-test game, a vertical scrolling obstacle avoidance care race.

7.3.3 Apparatus

For our experiment, we used Elemaya Visual Energy Tester (Figure 7.5), a simple portable tool to capture physiological signals. The tool was primarily designed for research into therapy using bio-feedback and medical treatments. Elemaya has 4 electroencephalography (EEG) channels, 2 electromyography (EMG) channels, and a channel for detecting galvanic skin response (GSR), heart rate (HR), and body temperature, respectively. In the experiment, we measured EEG, GSR, and HR as provided by Elemaya tool. According to 10-20 system (Niedermeyer and da Silva, 2004), Figure 7.6) EEG electrodes in Elemaya consist of 2 frontal (F1 and F2) and 2 temporal (T5 and T6) placed on a belt that the player wears on the scalp with two self-adhesive electrodes put behind the ears (reference electrodes)[4]. The electrodes require conductive gel on the player scalp and the sensors. On the other hand, the GSR sensor and photoplethysmo (HR) sensor were placed on the index finger and middle finger of the relaxed hand, respectively.

[4] http://www.elemaya.com/BrainMonitor.htm

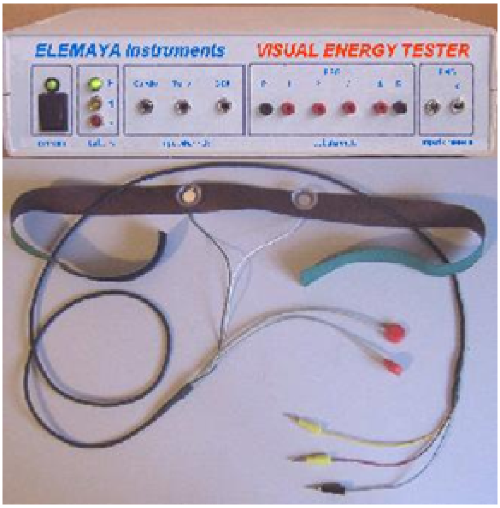


Figure 7.5: Elemaya Visual Energy Tester and its electrode belt.

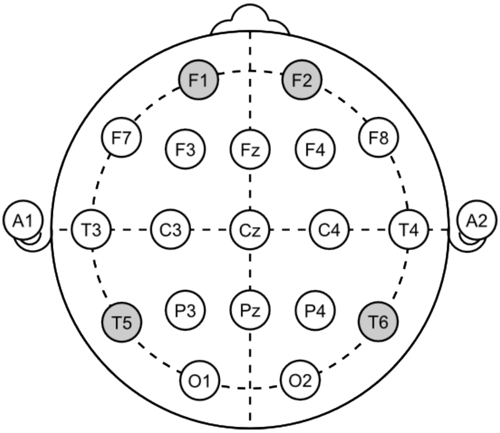


Figure 7.6: 10-20 system of EEG electrodes.[5]

7.3.4 Procedure

The experiment was a within-subjects design under three conditions: boredom, flow, and frustration. The tests were conducted in a laboratory and directly

[5]Creative Commons: http://creativecommons.org/licenses/by-sa/3.0/nl/deed.en_GB, author: Marius t́ Hart - http://www.beteredingen.nl

supervised by a researcher. This allows the researcher to observe the subjects' behaviors during the tests and to take notes if necessary. The procedure can be explained as follows.

1. The physiological sensors (EEG electrodes, GSR sensor, photoplethysmogram) are attached to corresponding body-parts of the subject.

2. The subject rests with closed eyes for 30 seconds before playing sessions to enable him/her to relax. This allows the subject to reset his/her physiological activities.

3. After hearing a beep, the subject read a very short introduction text about the goals and rules and starts to play the game with a given difficulty level for 3.75 minutes.

4. Upon completing a level, the subject fills out a questionnaire regarding his/her feeling during play.

5. Steps 2-4 are then repeated until the subject has completed all levels.

Our limitation was that we did not randomize the conditions. However, we argue that carryover effects were minimal since different conditions required different playing strategies, which diminished positive effects from practice. Moreover, each condition had enemies at random in number and positions. Conversely, the length of the whole sequence was less likely to overload the subjects nor to introduce fatigue effects.

7.3.5 Data pre-processing

Physiological signals were primarily used in clinical settings for therapies or medical treatments where the quality of signals are influenced by noises and the subject's physiological condition. Among those signals, the EEG signal is the most prone to noise coming from various sources, such as blinking, upper body movements, and neck contractions (Fatourechi et al., 2007). Since medical treatments mainly focus on capturing outliers of the brain activity that are relevant to diagnosis, it is critical for the patient to remain relaxed and stationary. Consequently, most clinical researchers perform several techniques to remove artifacts (Joyce et al., 2004), including artifact rejection using statistics and visual inspection. Some Matlab toolboxes, e.g. EEGLab[6] (Delorme and Makeig, 2004) and FieldTrip (Oostenveld et al., 2011), have provided such functionalities.

On the other hand, noises are normal and omnipresent in the gaming context since sometimes game play requires player to be physically active (Bos et al., 2010). Therefore, we should consider noises associated to body movement as

[6] http://sccn.ucsd.edu/eeglab/

part of the information that correlates with engagement (Bianchi-Berthouze et al., 2007).

In this experiment, all signals were sampled synchronously at 120 Hz (Elemaya sampling rate). Subsequently, Fast Fourier transform (FFT) was used to transform 120 samples of EEG, HR, and GSR signals into a 1-second epoch of power spectral density (PSD). This yielded 225 samples for 3.75 minutes of play in each condition. Initially, overlapping Hamming window were used to remove non zero values other than the intended frequencies (spectral leakage), but it does not affect the classification performance. Moreover, artifact rejection showed little significance in improving the EEG data (it affected only 5% of the total data). Consequently, the transformed data were immediately used for training and testing the classifier.

7.3.6 Features

Apart from the peripheral signals (GSR and HR), we have EEG signals that are subdivided into several wave bands: delta, theta, alpha, beta, and gamma (Chapter 6). However, the frequency ranges of each wave band are not homogeneously defined in literature primarily due to the differences in the domains of research and application.

In our experiment, the subdivision of EEG signals is presented in Table 7.1 in increasing order based on (Ang et al., 2012). The beta wave band is subdivided into three layers generally related to reasoning, alertness, and high mental activity (Babiloni et al., 2007). However, the gamma wave band is limited to 32 Hz due to Elemaya intrinsic limitation. In summary, we have 7 EEG wave bands at four electrodes and 2 peripheral signals (GSR and HR), i.e. 30 features in total.

Table 7.1: EEG wave bands

Band	Frequency
delta	1 – 4 Hz
theta	4 – 8 Hz
alpha	8 – 12 Hz
low beta	12 – 15 Hz
mid beta	15 – 20 Hz
hi beta	20 – 30 Hz
gamma	30 – 32 Hz

7.4 Exploratory data analysis

Prior to using predictive analysis, which is a classification technique in data mining, we performed exploratory and confirmatory data analyses on the transformed data.

7.4.1 Power spectral density (PSD) analysis

In this step, we evaluated the total PSD at four electrodes (F1, F2, T5, and T6 in Figure 7.6) across three conditions (with relaxing condition as a baseline) by averaging all epoch data on all subjects. PSD represents the power carried by the signal wave per unit frequency. Figure 7.7 shows that the relaxing condition has lower PSD magnitudes compared to the three gaming conditions. Moreover, it shows that the total PSD magnitudes tend to decrease with the increasing difficulty in T5 and T6 (temporal lobes). On the other hand, the PSD magnitudes dramatically drop during flow condition in F1 and F2 (frontal lobes). In addition, each electrode has distinct PSD shape and slopes which may convey different information and characteristics of different brain areas.

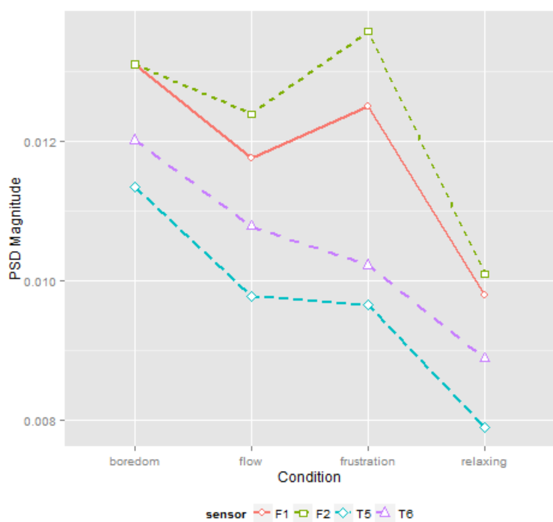


Figure 7.7: PSD at four electrodes (F1, F2, T5, and T6) given three conditions with relaxing condition as a baseline.

Subsequently, we evaluated the PSD at four electrodes in each EEG wave band across three conditions. The results show that the delta wave dominates the PSD values, probably due to eye movements and blinking, whereas alpha and low beta waves are the largest at frontal lobes under the frustration condition (Figure 7.8). The boredom and the flow conditions tend to overlap more in the frontal areas, while the flow and the frustration conditions tend to overlap more in rear temporal areas.

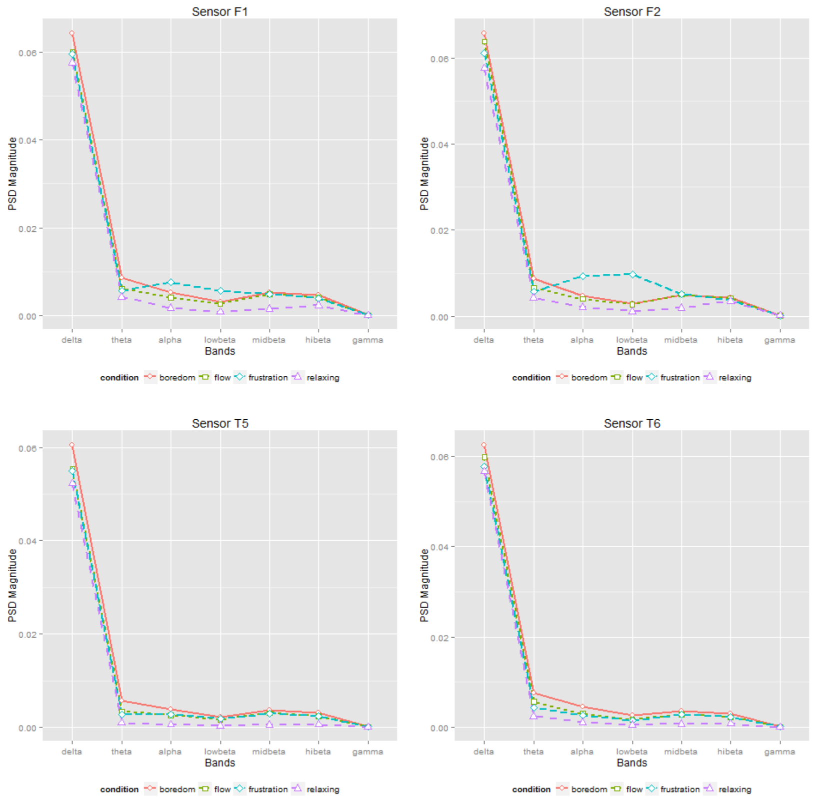


Figure 7.8: PSD of each EEG wave band at four electrodes (F1, F2, T5, and T6) given three conditions with relaxing condition as a baseline.

In addition, we evaluated the PSD changes of each EEG wave band at four electrodes and the total PSD changes across all conditions (Chanel et al., 2011). Theta wave tends to attenuate with the increasing difficulty, while alpha and low beta waves tend to attenuate in the flow condition (Figure 7.9). Pellouchoud et al. (1999) found that alpha wave attenuates from resting to game-playing condition. Moreover, Sheikholeslami et al. (2007) investigated theta and alpha waves during 50 minutes game-playing and found that alpha wave progressively attenuated. Therefore, we may infer that the subjects entered flow. The subjects also confirmed after the experiment that they experienced flow. On the other hand, mid beta, hi beta, gamma waves only attenuate at T6 during flow.

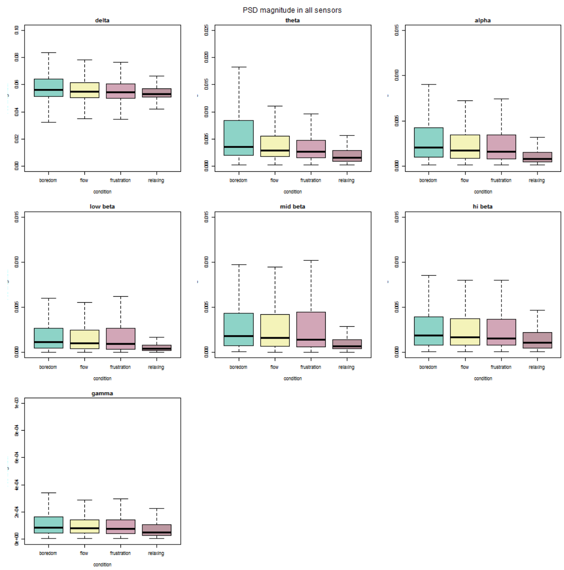


Figure 7.9: Total PSD changes of each EEG band averaged.

7.4.2 Analysis of variance (ANOVA) of the PSDs

After conducting PSD analysis, we performed 32 separate one-way ANOVAs on 30 data features and the average PSD as dependent variables and conditions as factor, to investigate different characteristics of the features. Table 7.2 presents the results of one-way ANOVAs grouped by each sensors which shows mid beta wave at T5, hr, and gsr are not statistically different (red-colored rows). This indicates that those features are not of interest to be used in predictive analysis.

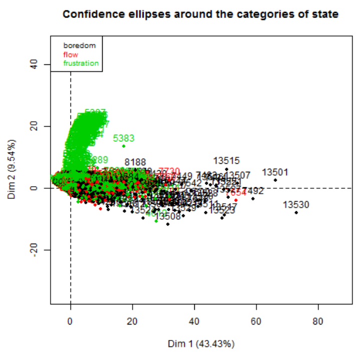
Subsequently, post-hoc Tukey's HSD (honest significant difference) tests were carried out to find the differences between conditions, i.e. boredom (C1), flow (C2), and frustration (C3). Theta wave at right hemisphere (F2, T6), alpha wave at F1, and average PSD at F1, F2, T6 are useful to identify different conditions (Table 7.2). All EEG waves at T5 are significant only at differentiating flow (C2) to frustration (C3). The post-hoc analysis also shows that flow and frustration are most likely discernible with 25 features followed by boredom and flow with 10 features, and boredom and frustration with only 8 features.

Table 7.2: One-way ANOVA $F(214, 847)$ and Tukey HSD on physiological signals.

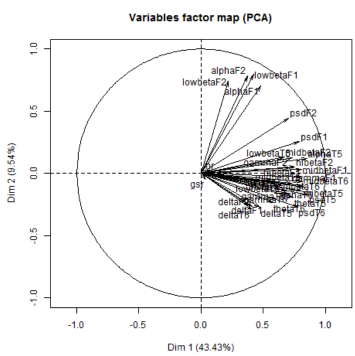
Sensor	Feature	F-value	p-value	C1-C2	C1-C3	C2-C3
F1	delta	36.24	< 0.0001	-	-	delta
	theta	82.96	< 0.0001	-	-	theta
	alpha	47.16	< 0.0001	alpha	alpha	alpha
	low beta	91.34	< 0.0001	low beta	-	-
	mid beta	6.44	< 0.01	-	-	mid beta
	hi beta	22.31	< 0.0001	-	-	hi beta
	gamma	33.14	< 0.0001	-	-	gamma
	avg psd	36.27	< 0.0001	avg psd	avg psd	avg psd
F2	delta	24.10	< 0.0001	-	delta	-
	theta	76.78	< 0.0001	theta	theta	theta
	alpha	66.66	< 0.0001	alpha	-	-
	low beta	117.85	< 0.0001	low beta	-	-
	mid beta	0.21	> 0.5	-	-	-
	hi beta	9.38	< 0.0001	hi beta	-	-
	gamma	6.58	< 0.01	-	-	gamma
	avg psd	18.66	< 0.0001	avg psd	avg psd	avg psd
T5	delta	87.94	< 0.0001	-	-	delta
	theta	127.49	< 0.0001	-	-	theta
	alpha	38.28	< 0.0001	-	-	alpha
	low beta	19.33	< 0.0001	-	-	low beta
	mid beta	14.00	< 0.0001	-	-	mid beta
	hi beta	29.55	< 0.0001	-	-	hi beta
	gamma	48.66	< 0.0001	-	-	gamma
	avg psd	98.07	< 0.0001	-	-	avg psd
T6	delta	39.26	< 0.0001	-	delta	-
	theta	75.86	< 0.0001	theta	theta	theta
	alpha	55.63	< 0.0001	-	-	alpha
	low beta	52.30	< 0.0001	-	-	low beta
	mid beta	16.91	< 0.0001	-	-	mid beta
	hi beta	28.50	< 0.0001	-	-	hi beta
	gamma	41.90	< 0.0001	-	-	gamma
	avg psd	74.66	< 0.0001	avg psd	avg psd	avg psd
HR	hr	0.37	> 0.1	-	-	-
GSR	gsr	0.42	> 0.1	-	-	-

7.4.3 Relevant features

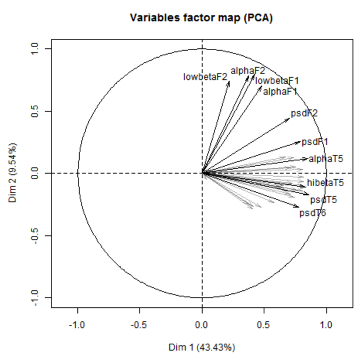
Feature selection is very important in classification, since some features might actually act as noises that disrupt the performance of classifiers. Thus, we compared the performance of classifiers which trained using different set of features. For instance, we used 31 significant features from one-way ANOVAs (non-colored features in Table 7.2), all available features (34 features), and other different combinations of features to obtain better classifier.



(a) Confidence ellipse of each condition.



(b) Two principal components given all features



(c) Top-ten features contributed to the principal components

Figure 7.10: PCA Analysis for EEG, GSR, and HR.

To this end, we conducted principal component analysis (PCA) to translate all the features into principal components, i.e. components that highly contribute to the data variability. In principal, PCA performed orthogonal transformation to convert a set of possibly correlated features into a set of new linearly uncorrelated features (Abdi and Williams, 2010). The results show that boredom and frustration conditions tend to differ, whereas the flow condition heavily overlaps with the other two conditions (Figure 7.10(a)). This gives us an preliminary insight that frustration is easier to recognize compared to the other two conditions. Top-ten features contributed to the data variability include alpha in F1, F2, T5; low beta in F1 and F2; hi beta in T5; and average PSD in all electrodes (Figure 7.10(c)).

Subsequently, we performed an *n* factor analysis to the principal components to obtain *n* optimal principal components for classification (Raîche et al., 2013). The results show the first 7 principal components are in the optimal region (Figure 7.11). Thus, we built one of the classifiers using 7 principal components as comparison to using features.

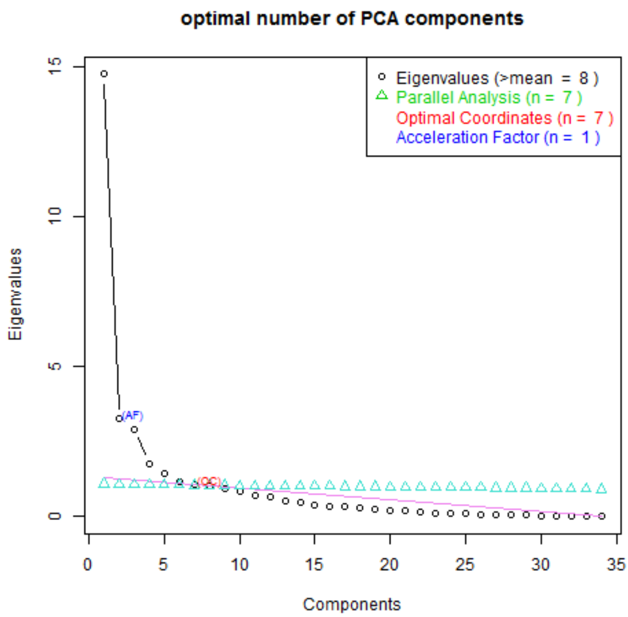


Figure 7.11: Optimal number of PCA components.

Observing the first component coefficients, the most important original features are: alpha, low beta, mid beta, and high beta, for all sensors. This confirms our expectations on the PSD values of the various bands for the various levels (Figure 7.8).

7.5 Predictive analysis

In this step, classification techniques in machine learning were used to build classifiers for detecting flow. Consequently, the decision made by the classifiers could be used automatically for difficulty scaling by, for instance, adjusting the level of challenges in the game. The power of each classifier was determined by the prediction accuracy and the confusion matrices for three conditions. There were two supervised learning approaches used in our work: non-probabilistic approach using support vector machine (SVM), and probabilistic approach using Bayesian framework. SVM was chosen since it has been very successful for pattern recognition shown by remarkable experimental results in very diverse domains of application (Blanchard et al., 2008), in particular EEG classifications (Garrett et al., 2003; Rakotomamonjy and Guigue, 2008; Sun et al., 2010a). On the other hand, simple statistical procedures proved to be very competitive for classification and mostly produced good "out of the box" results without the inconvenience of delicate and computationally expensive hyperparameter tuning such as SVM (Meyer et al., 2003). Thus, we also explored three naive Bayes algorithms which have more relaxed structures compared to conventional naive Bayes: Tree Augmented Network (TAN) (Friedman et al., 1997), Hidden Naive Bayes (HNB) (Jiang et al., 2009), and Weightily Averaged One-Dependence Estimators (WAODE) (Jiang and Zhang, 2006).

7.5.1 Collective classifier for all subjects

In this step, we created several user independent classifiers using data from all subjects. Five feature settings were used to investigate the performances of the classifiers trained using different features. Moreover, this allowed us to investigate the effect of feature selection and/or transformation, such as ANOVA and PCA mentioned before, towards classification of physiological signals. The five feature settings are as follows.

1. 34 features (34F): 28 BW, 4 average PSD (APSD), 1 HR, and 1 GSR
2. 31 features (31F) from ANOVA (Table 7.2)
3. 30 features (30F): 28 BW, 1 GSR, 1 HR
4. 28 features (28F) from BW only
5. 7 components of PCA of Figure 7.10(c) (7PCA)

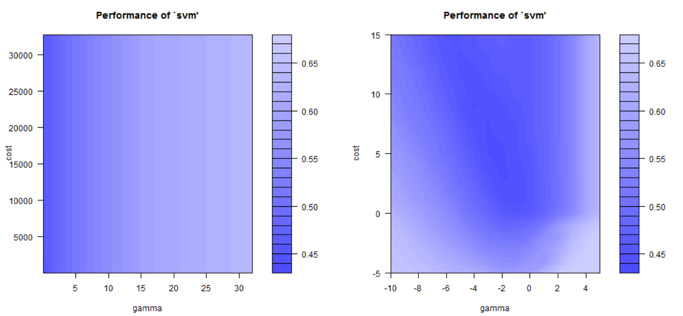
All of the features were normalized into z-score prior to classification ($\mu = 0, \sigma = 1$) and two configurations were used for selecting the training set and test set:

1. 66% data of each subject in each state as a training set while the rest (34%) was used as a test set, which was subsequently encoded as 66-34.

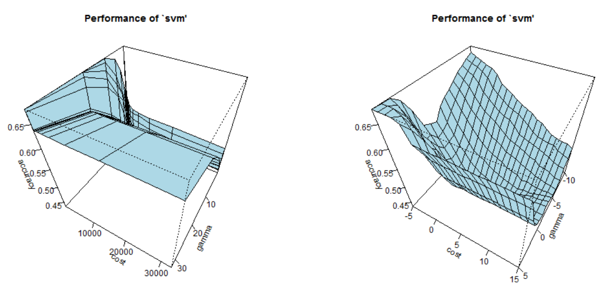
2. A training set from 15 subjects (10,125 samples), and a test set from 7 subjects (4,725 samples), which was subsequently encoded as 15-7.

The objective of using two different sets was to test whether subjects whose data are completely unseen (i.e., not used for training the classifier) affect the performance of the classifier.

For SVM classification, we used LibSVM[7], state-of-the-art library for support vector machine implementation, in R (package *e1071*) for SVM classification (Chang and Lin, 2011; Hornik et al., 2006). For model selection, RBF kernel was used for practical reasons (Hsu et al., 2003). Grid-search methods were performed to obtain near optimal hyperparameters with $C \in 2^{(-5,15)}$ and $\gamma \in 2^{(-10,5)}$ and k-fold cross validation was used for training the classifier with $k = 6$ (Hsu et al., 2003).



(a) Hyperparameter search space in 2D. (b) Hyperparameter search space in 2D ($\log_2$)



(c) Hyperparameter search space in 3D (d) Hyperparameter search space in 3D ($\log_2$)

Figure 7.12: Tuning SVM hyperparameters.

[7] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Figure 7.12(a) and 7.12(b) show that the classifier has lower errors at $\gamma \in (0.1, 4) \approx (2^{-3}, 2^{2})$ and $C \in (1, 100) \approx (2^{0}, 2^{7})$. Moreover, the error surface - as measure of accuracy - decreases as γ decreases and reaches the lowest at approximately 2^{-2} (Figure 7.12(d)). In contrast, the error surface decreases with the increasing C and reaches the lowest at approximately 2^{5}. Hence, The grid-search using k-fold cross validation yielded the best hyperparameters for the given ranges are $C \approx 32$ or (2^{5}) and $\gamma \approx 0.25$ or 2^{-2}. Subsequently, we used those hyperparameters to train several SVM classifiers using feature settings and data configurations above; and predict the samples in the test sets. The results are presented in Table 7.3.

Table 7.3: Performance of SVM collective classifiers.

data set	features	accuracy (%)	σ	min	max
66-34	34F (BW, APSD, GSR, HR)	**49.6**	7.6	35.1	**66.2**
	31F (BW - midbetaF2, APSD)	49.2	7.6	**36.8**	63.6
	30F (BW, GSR, HR)	49.0	7.9	34.6	**66.2**
	28F (BW)	48.4	8.2	36.0	63.2
	7PCA (PCAs)	41.8	7.0	26.3	60.9
15-7	34F (BW, APSD, GSR, HR)	36.9	6.7	32.1	51.4
	31F (BW - midbetaF2, APSD)	**37.6**	3.6	**33.5**	41.8
	30F (BW, GSR, HR)	36.8	7.2	31.2	**52.0**
	28F (BW)	37.0	3.9	32.1	41.5
	7PCA (PCAs)	36.5	6.1	29.6	45.8

Results of SVM collective classifier - expressed in terms of recognition rates, where the last two columns show the average recognition rate of the worst and of the best users, respectively (Table 7.3). In 66-34 setting, 34F provides the best classifier performance and closely followed by 31F and 30F. On the other hand, 31F provides the best results in 15-7 setting followed by 28F. In addition, 31F shows a stable performance indicated by lower standard deviations. This means feature selection using one-way ANOVA better handles the unseen data. GSR and HR marginally contribute to the classifier performance in 66-34 setting, while they actually decrease the classifier performance in 15-7 setting. From the confusion matrix of 34F (Table 7.4), it can be seen that frustration (C3) and boredom (C1) were better classified compared to flow (C2).

Table 7.4: Confusion matrix of 34F in SVM collective classifier (in %).

	boredom	flow	frustration
boredom	68.7	20.1	11.2
flow	41.0	37.8	21.2
frustation	36.5	21.2	42.3

On the other hand, after training and testing several user independent naive Bayes classifiers with HBN, TAN, and WAODE structures using WEKA[8], we obtained the performances of the Bayesian classifiers to be slightly lower with respect to SVM classifiers. For 30F setting, the average accuracies of Bayesian classifiers were 45.3% for HNB, 44.3% for TAN and 46.2% for WAODE. Since the prediction step in both SVM and Bayesian classifiers produced probabilities for all the conditions, we fused the decision from SVM and Bayesian classifiers using majority voting. For instance, the SVM classifier produced probabilities for the first sample **0.42 C1**, 0.28 C2, 0.3 C3, whereas the Bayesian classifier produced 0.32 C1, 0.2 C2, **0.48 C3**. Then, fusing both classifiers yields 0.37 C1, 0.24 C2, **0.39 C3**. This method improves the predictive performance for collective classifier by ±1.5% (50.1%, σ = 8%, min =29.8%n,max =64.1%).

7.5.2 Individual classifier for each subject

Individual (user dependent) classifiers differ from collective classifiers since each of the individual classifiers were tuned specifically for one subject. In other words, we developed 22 different classifiers for 22 subjects, *ergo*, we were only able to use 66-34 configuration for this purpose. This also implies we need to train each individual classifier (22 times) but with less training sets for each classifier (i.e. 447 samples). Table 7.5 expressed the performance of SVM individual classifier in terms of average recognition rates of all classifiers, are reported in , where the last two columns show the average recognition rate of the worst user and of the best user.

[8] http://www.cs.waikato.ac.nz/ml/weka/

Table 7.5: Performance of SVM individual classifiers in 66-34 setting.

features	accuracy (%)	σ	min	max
34F (BW, APSD, GSR, HR)	59.2	10.0	43.4	83.8
31F (BW - midbetaF2, APSD)	**61.0**	13.3	44.7	**96.0**
30F (BW, GSR, HR)	59.6	10.7	42.5	85.1
28F (BW)	60.8	11.3	**46.0**	86.4
7PCA (PCAs)	45.9	11.6	31.6	74.6

31F has the highest accuracy and is tightly followed by 28F. Results for 34F and 30F are slightly worse (lines 1 and 3 of Table 7.5). This shows that EEG measures are useful to predict the current state of the subject while GSR and HR are less reliable to be used as features. On the other hand, we observed that adding different features, such as the *attention ratio* (Putman et al., 2010) and the total electrode power, does not improve results. This is because attention ratio were positively correlated to beta and negatively correlated to theta. The confusion matrix of 31F in SVM individual classifiers is reported in Table 7.6. Similar to the collective classifier (Table 7.4), it can be seen from Table 7.6 that frustration (C3) and boredom (C1) are well classified, while flow (C2) classification is less accurate.

Table 7.6: Confusion matrix of 31F in SVM individual classifier (in %).

	boredom	flow	frustration
boredom	63.5	24.6	11.9
flow	17.9	49.8	32.3
frustation	11.7	18.6	69.7

Feature selection that may cut off unsubstantial features - using one-way ANOVA - produced better classifiers performance for collective and individual classifiers. This emphasizes the importance of feature selection in physiological signals. Our finding differs to Sun et al. (2010b) where they considered all the original physiological data are important in classifying motor imagery tasks. Dimensionality reduction using PCA produced considerable loss of performance, i.e 2-7%, but the number of features is reduced to one quarter of the original, which could be important for a prospective implementation of a real-time validation and training system with respect to the training time.

The SVM execution times were as follows. For model selection and training, individual classifiers require 2 minutes, while the collective classifier requires approximately 12 hours. All the times were taken on an Intel Xeon CPU E5-2620 v2 @2.1 GHz (12 CPUs) processor, 16-GB RAM, Windows 7 Pro 64 bits.

From our analysis, we have seen that some bands are more significant (alpha and beta) to others, but all the electrodes provide useful information. Subsequently, we explored the performance of the electrodes in pairs which represent frontal, left, right, and temporal hemispheres of the scalp. For each pair, we used all the brainwave bands (14 features). Table 7.7 shows the results that pair of electrodes are able to provide significant information. Although evidence in the literature mentioned the differences between two cerebral hemispheres in morphological, biochemical, and functional characteristics (Steinmetz et al., 1991), we could not observe significant differences among our subjects (five left-handed, 17 right-handed).

Table 7.7: Performance of SVM in four scalp regions.

pair	accuracy (%)	σ	sensors
right	52.8	12.6	F2 and T6
left	56.6	10.1	F1 and T5
frontal	52.5	13.8	F1 and F2
temporal	56.4	11.9	T5 and T6

Using Bayesian network (BN), the results from all test users in 28F are lower than in the SVM (HNB: 56.6%; TAN: 56.5;WAODE: 56.2%). Compared to collective classifier, the differences are substantial most likely due to the fact that BN exploit inherent probabilities in the data. This made the BN performed less in individual classifiers given less number of samples. In addition, collective classifier offers an advantage to individual classifiers in the sense that it requires training one model for all subjects. Nonetheless, the data collection times (3.75 minutes for each condition) and the lower classifier training time opt for individual training which make the latter feasible even in consumer applications.

7.6 Task comparison

To define and analyze the spectral peculiarities of game playing, with respect to other intellectual tasks, we selected five (four males and one female) among 22 subjects to perform three different tasks in different experimental sessions. The first task required the subjects to sequentially relax for 3.75 minutes, rest for 30 seconds, and read an Internet journal article which is considered as easy reading for equal duration. Subsequently, the second task began with resting for 30-s and ended with solving three mathematical quizzes for 3.75 minutes.

Figure 7.13 shows the average PSD of the different bands for the three tasks compared to gaming in the flow condition from the previous experiment. The

PSDs were obtained through FFT in 1 second windows as previously done. It shows that the flow condition tends to have the energy distributed evenly across different frequencies.

Three binary SVM classifiers with 28 features were trained to identify whether the flow condition could be distinguished from other tasks. The results are largely positive, with average accuracy of 86% (vs. relax condition), 87.9% (vs. reading task), 84.1% (vs. math problem solving). This can be attributed to the fact that binary classifiers are easier to construct compared to multiclass classifiers since less number of samples resided in the margin of the support vectors. Moreover, the samples turned out to be linearly separable for each pair.

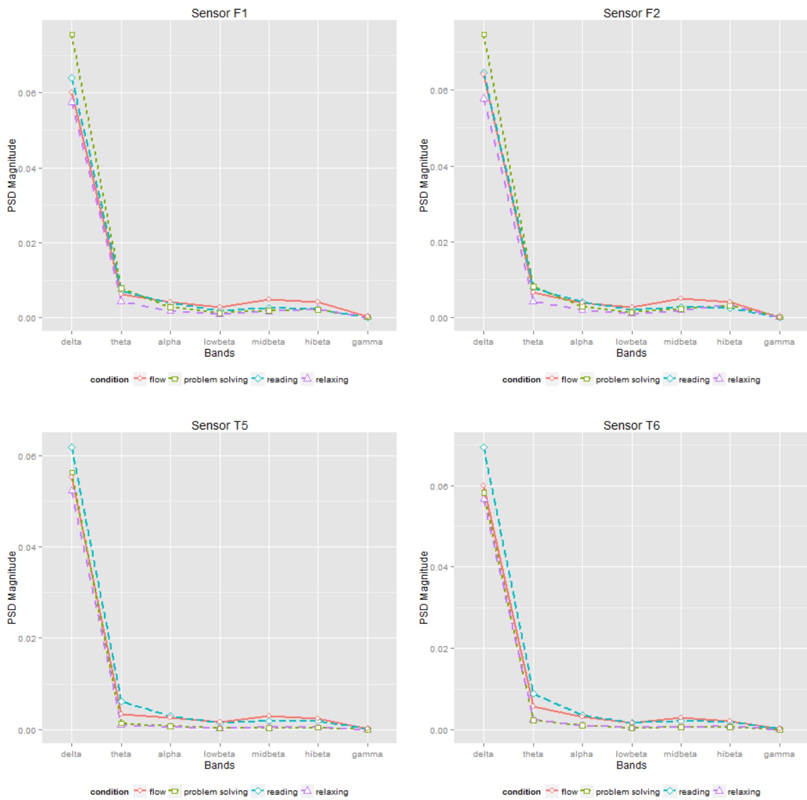


Figure 7.13: PSD of flow condition vs other intellectual tasks

7.7 Discussion and conclusions

A similar video-game spectral analysis was performed by Pellouchoud et al. (1999) but in a narrower frequency ranges (6-13 Hz band) and focused on mental workload. Research conducted by Chanel et al. (2011) showed that, after fusing of the two signal categories: EEG and peripheral signals for a total of 77 features, the collective classifier achieved accuracy up to 63%. This result was computed from samples of 300 seconds width. However, the classifier performance constantly decreased when the sampling window reduced to 30 seconds width with accuracy of 51% and 46% for EEG and peripheral signals, respectively. Consequently, this setting is not very suitable for real-time analysis and feedback, which is required in dynamic activities such as gaming. Moreover, predictive algorithms typically require significant numbers of samples which justify our choice for 1 second width windows. This is common in clinical research and workload assessment (Chaouachi et al., 2011) although it is less effective for peripheral signals.

This is the first, that is known, result on 1 second window width for video-game analysis is reported. Liu et al. (2009) reported that using non neurophysiological signals significantly contributed to better classifier performance (15% higher to Chanel et al. (2011)). However, their system is user dependent and requires six 1-h training game sessions for each participant, while Chanel et al. (2011) recorded 1 minute baseline for each new subject. In our work, we use both generic classifier (collective) such as Chanel et al. (2011) and individual training such as Liu et al. (2009). It took 4-5 minutes of training time to train individual classifier. Hence, it is feasible to train an individual classifier for real-time feedback with 61%.

Although EEG sensors are not a comfortable tool for the user due to the gel fixing of the electrodes, which limit their performances, technological advances in human-computer interaction are likely to fill this gap with appropriate devices (Gürkök and Nijholt, 2012; Van de Laar et al., 2013), in particular with dry sensors, that are more practical but still feature lower performance (Van Erp et al., 2012). Based on our results, the most informative bands are those around alpha and low beta for distinguishing states in gaming (Figure 7.8), while mid-beta for discriminating gaming from other tasks (Figure 7.13). A similar video game spectral analysis was performed by Pellouchoud et al. (1999), but this only included the 6-13 Hz band and focused on mental workload, not on flow. Peripheral signals add little information in detecting state. Classification of three conditions of user states yielded moderate accuracy, in particular using SVM. Bayesian frameworks produced lower accuracy with limited advantages in one case of information fusion. Flow is more difficult to characterize compared to boredom and frustration while feature selection using one-way ANOVA provides more stable and better classifier performances. A personalized system could be easily implemented in a consumer context as a service using individual classifier given that the data collection and training time is reasonable.

7.8 Future works

Our work provided a building block for supporting the educator in easy and effective serious game development, in particular the use of adaptivity as a service in games. Since adaptivity requires many modules, this work supplies additional information in terms of user state to the adaptation module. For instance, if the physiological state indicates frustration and the user performance (e.g., remaining lives, number of correctly solved puzzle) drops, the game should present less difficult enemies or different types of feedback. Thus, an adaptivity mechanism should be further devised using the classifiers mentioned above. In addition, continually adapting the challenges may keep players in "the zone" (flow), which in turn may improve the feeling of learning at the very least. This is essential in particular to improve students' motivation in learning.

In classification perspective, this work also enables further studies in investigating new methods to improve classification performance. In this regard, It would be interesting to investigate different feature extraction and transformation techniques. Nicolas-Alonso and Gomez-Gil (2012) suggests a wavelet analysis, while Krusienski et al. (2011) highlights the potential of recurrent artificial neural networks to represent complex, nonlinear spatio-temporal patterns that are not captured by the current approaches based on PSDs. Complementary information such as amplitude and phase may contribute to a better representation. Common spatial pattern (CSP), data analysis and a classification technique based on spatial filtering, is worth to consider as it reduces the effect of volume conduction of EEG multichannel signals (Blankertz et al., 2008). Furthermore, We did not perform artifact rejection on the EEG, which possibly degraded the performance of the classifiers. It would be useful to recognize artifact in the EEG data and subsequently perform artifact rejection. One of the possible ways is to visually record a player's gaming activities, and thoroughly inspect the EEG data for artifacts with the aid of the visual recording. Thus, we could annotate contaminations in the EEG data at certain windows and, subsequently, build an automatic classifier for recognizing artifacts prior to feeding the EEG data into either individual or collective classifiers. However, this approach may be not suitable for a real-time classification.

An application field that we would like to address concerns difficulty scaling in serious gaming, i.e. to keep a user in flow by, for instance, increasing challenges when a player is detected as becoming bored. This implies the added value of retaining player interest to educational content for longer periods. This requires investigating the neural correlates between entertainment, in particular flow, and learning.

Difficulty scaling in this approach requires research to integrate non-invasive BCIs technology with existing gaming hardware and software (Van Erp et al., 2012). This allows the implementation of a real-time state monitoring system, including standard hardware for signal acquisition and software for processing embedded in the game engine or as a callable service.

Part IV

Conclusions

Chapter 8

Conclusions

"I know one thing, that I know nothing." - Socrates, paraphrased from Plato's Apology.

8.1 Main conclusions

Our research was driven by the need of Serious Games Society (SGS) for efficient serious games (SG) development under service oriented architecture (SOA) platform. This will benefit researchers and practitioners in the area of serious games since this will prevent them from reinventing the wheel by using well established and ready-to-use services for their games, and promote easy game authoring. We focused on two aspects of efficient SG development: the format and the delivery strategy.

The format, and the evaluation of flow & learning

The first aspect includes the extensible game format, the architecture, and the game features. We evaluated the game in terms of user perception and user performance to improve the game features as services. Furthermore, we evaluated the effect of a game feature (i.e., tutoring tool) on flow and learning, and investigated the relationship between flow and learning. All of these are valuable to provide ready-to-use services, in particular what services to implement (e.g., game features such as tutoring tool) and how to implement and use the services (e.g., tutoring tool as a learning assistant). Furthermore, our experiments indirectly set up a guideline to evaluate services in terms of flow and learning. This is particularly useful for evaluating the effectiveness of game features and games, which in turn improve the game features as services and the games as a whole.

Therefore, in Chapter 3, we created a game format and subsequently developed a serious game for learning physics, specifically classical mechanics based on a specific learning objective. We infused simulation to allow players experiencing phenomena, puzzles with immediate feedback to encourage reflection and conceptualization, and successive interrelated tasks and simulation to allow experimentation. We also implemented a tutoring tool for scaffolding. We tested the game for its usefulness with 10 participants. The results show that the participants perceived the game to be educative and moderately entertaining. However, presenting puzzles limited the exploratory behavior of the participants to play with the simulations, since they were more concerned with solving the puzzles and only perceived specific events in simulation related to the puzzle. Users also preferred to have all game artefacts available all the times regardless they were appropriate for the task at hand. Although this seem contradictory to working memory limitation, this finding is similar to 'just-in-case' presentation of information in serious games (Van der Spek, 2011). This may relate to providing a higher degree of control to players as suggested in the flow antecedents. We concluded that sense of control and clear goal are essential in designing and developing serious games, considering flow. The work in Chapter 3 extends the works in (Bellotti et al., 2009b, 2010) by providing an extensible game format and services to support pedagogical authors in providing games for learning, in particular for easy and effective game creation. Moreover, the experiment intended to get a better grip on how and if flow is used for game development improves learning, and thereby leads to better learning games. This work answers the first part of our research questions (i.e., *what services need to be implemented and how to implement the services?*).

Subsequently, using improved prototype of Chapter 3, in Chapter 5, we evaluate the effectiveness of physics game in terms of flow and the learning outcomes from the target users, in particular how useful the tutoring tool as a service in physics games for improving the learning outcomes and retaining flow. We adopted eGameFlow questionnaire (Fu et al., 2009) to measure flow (Chapter 4) and constructed a test set to assess learning outcomes in the form of knowledge and misconceptions, and procedural learning. Subsequently, we compared two versions of the physics game: with the tutor and without the tutor. We hypothesized that the tutoring tool would have disrupted flow but supported learning. Surprisingly, we found that the one with tutoring tool received significantly higher level of flow compared to the one without tutor but no significant effect on the learning outcomes, although, the one with tutoring tool had slightly better learning outcomes. Most likely that learning companion, i.e. the tutoring tool, improve the motivation for playing and learning with the game. (Woolf et al., 2010) found a similar effect of a learning companion for low achieving students and students with disabilities. However, in our case, our participants were quite knowledgeable in physics. This means that they considered the learning companion to have added meaningful knowledge to their understanding. We also found that subjective learning, i.e. feeling of learning, was align to flow. However, it did not correlate with the learning outcomes. The work in Chapter 5 advances the work in (Kiili, 2006; Fu et al., 2009) by evaluating games in terms of both flow and the learning outcomes, and providing a reproducible procedure for evaluating the effectiveness of both game features and games in terms of flow and learning. Furthermore, this supports the educators in effective game creation, in particular how a tutoring system could be used to obviate the need for active guidance and inquiry stimulation by the teacher. This work answers the second part of our research questions (i.e., *How we evaluate game features (services) in terms of flow and learning? Is there any relationship between flow and learning?*).

Based on our findings, we concluded that creating services for serious games to support experiential learning is along process which involves designing, developing, and evaluating the usefulness of the services in terms of flow and learning. Moreover, standardizing the evaluation of games and game features as services could benefit researchers in serious games. This includes both the tools and the procedure for the assessment.

The adaptivity

In the second aspect, we investigated the use of physiological signals for real-time adaptivity in games, since physiological signals represent functions of experience (Fenz and Epstein, 1967). We divided user affective state into three classes: boredom, flow, and frustration, and tailored a game into three levels accordingly. For each class, we recorded 3.75 minutes of 22 participants' physiological data (EEG, HR, and GSR) while they were playing the game (Chapter 4). Subsequently, we sliced the physiological data into 1-s window frame and performed two classification algorithms, SVM and Bayesian framework, to classify players' states. The results show that EEG, HR, and GSR were feasible to be used for real-time adaptation with 61% and 50.1% accuracy for user dependent and user independent classifiers, respectively. Based on the results, we recommend the use of individual classifiers due to less training time and higher accuracy. This approach supports the educator in easy and effective game creation in the sense that it supports automatized testing and game balancing to improve learning, so that the educator doesn't have to determine the proficiency level of every student and balance the game accordingly. This answers the last part of our research questions (i.e., *Can we develop adaptivity as a service by using physiological signals?*).

8.2 Insights to implement and improve services for efficient SG development

As a result of the research outlined in this dissertation, we summarized several practical insights for creating services for efficient SG development as follows.

1. Creating a format provides extensibility and reusability in authoring/customizing tasks in games. This also aid teachers to author/customize game tasks by themselves. To this end, component-based and, overall, service oriented architecture are key points to support easy game and game task creation since game authors just simply reuse/call the services to implement certain functionalities in their games. One of

examples is to apply scalable techniques to create services, such as natural language processing in the tutoring tool (machine learning). This enables pedagogical authors to scale up the knowledge of the tutor easily.

2. Appropriate solutions can create new SG mechanics that are able to join goals that in several cases are conflicting. This was shown by the learning companion in games which turned out non-intrusive to flow. One of ways is by integrating the learning companion as part of the interaction in games.

3. eGameFlow questionnaire accompanied by test set could be very practical for serious game and game features evaluation. They also provide immediate feedback for game designers in improving the game features and the game as a whole.

4. It is more effective and efficient to use individual physiological signals to train adaptivity services compared to a collective approach. This was shown by significantly lower training time and higher classification accuracy. In addition, an individual approach paved the way for real-time adaptivity.

5. Adaptivity based on physiological signals should be mainly used by a game for automatic adaptation of parameters, such as speed, score, music (and also at certain extent, difficulty of tasks), since it was more of a neuro-characterization than the game features for keeping the user in flow. Game features (e.g., the tutor or others) should be built atop of it (i.e., exploiting real-time information from it).

8.3 General limitations of the research

To extend the SG architecture, we created services to support SG development in which we considered extensibility, scalability, and reusability. As an example of using the services, we created a game for learning physics. This laid the important groundwork, but the development of the authoring tool -to test the whole system in the perspective of the game authors- lies outside the scope of the research. This was because primarily due to one of our main goals, i.e. evaluating flow and learning of the game and game features to improve the implementation of the services. In addition, the tutoring tool was designed specifically for the physics game with a very specific learning objective. Although, we have performed a controlled experiment and the tutoring tool worked for our case in terms of flow and learning, generalization to other game settings, game types and learning goals should be performed with caution (Shapiro and Peña, 2009). Therefore, we also researched a more generally applicable (though less convenient) flow adaptation.

In our work, relatively short durations for experiments might affect the outcomes. First, for the physics game, we provided the participants with 20 minutes to play the game, a relatively short duration for learning. However, this was necessary since the total time needed from pre-test to post-test was approximately 1 hour. A longer duration for the experiment would also have compromised the participants due to exhaustion. Second, for the adaptivity, we recorded 3.45 minutes of physiological signals for each game difficulty level (total of 10.35 minutes for three difficulty levels for each participant). This might be inadequate to perfectly render the states (boredom, flow, frustration). However, in a consumer context, the time -3 minutes data collection and 1-2 minutes user dependent classifier training- is reasonable.

Based on our results, flow improves the perceived performance, but not the actual performance. However, the limitations could be that the game was too short (as already mentioned), the material was too difficult or too easy, the game focused on problem solving while the knowledge (and the test) is perhaps more declarative, and we didn't measure the effect of flow on learning attitude and cognitive load/effort. Furthermore, we performed a post-hoc evaluation of flow; it is unclear whether flow was directly associated with learning, or whether flow levels were simply not high enough to see a real effect on learning, which in turns it needs to be stimulated more.

8.4 Implications and future research

In creating services to support SG development, we found that the tutoring tool was useful in our scenario but we could not generalize the results in different settings. For instance, it would be interesting to investigate the effect of the tutoring tool in totally different setting, such as in Remission game, an action based serious games for improving patients' knowledge about cancer (Beale et al., 2007). Moreover, our participants were equal in terms of knowledge and skills, and thus, it would be useful to know how players with higher skills perceive the learning companion (the tutoring tool) compared to low achievers (Woolf et al., 2010). As learners achieve mastery and gain competence, the role of scaffolds, including feedback and tutoring tools, need to be diminished. Consequently, future research needs to investigate the hints and the pace for diminishing the scaffolds. On the other hand, we could also extend the physics game to address different topics in physics based on a well defined curriculum. This would allow researchers to perform longitudinal research into the efficacy of the physics games and the tutoring tool towards students' motivation and learning outcomes.

In the light of design, our approach could be reproduced for different game settings, game features (services), game genres, and learning objectives. Moreover, this would provide more information in the evaluation games and game features in terms of flow and learning. For instance, the benefit of social interaction towards both flow and learning outcomes. Moreover, we could also arrange the dimensions of flow in order of importance, according to both characteristics of the games and the players. Different game genres may have different prominent features. In addition, players may have different preference and playing style, even for a single game (Kline and Arlidge, 2003). This could improve the game design by accommodating different playing styles, which in turn opens up new avenue of research into playing style adaptation.

Our experiments did not clearly provide the relationship between flow and learning. Intensively reproducing our approach with different games and different features could provide not only an illustration of the relationship between both, but also design practice of useful features and services in particular game genres. This in turn will provide Serious Games Society not only with a catalogue of services but also useful services for certain types of game (grouping of functionalities). Graesser et al. (2009) argued that learning should be painful to promote deep learning. In contrast, Kiili (2006) found positive correlation between flow and learning. Therefore, future research could employ our method to investigate the role of flow in serious games for different levels of learning (surface learning and deep learning). This would further clarify the correlation between flow and different levels of learning.

In the adaptivity as a service part, we showed that brainwaves have potential to capture flow in games for a real-time adaptation. Certainly, EEG is not a comfortable tool for the user, in particular because of the gel fixing of the electrodes. Thus, both measurements and usability could be degraded. However, technological advances are likely to fill this gap with appropriate devices (Obbink et al., 2012), in particular with more practical dry sensors, though with lower performance (Van Erp et al., 2012). The potential of real-time adaptation also opens up new avenues of online incremental learning. Moreover, using the physiological signals for adaptivity will enable researchers to gain more information regarding what the physiological signals portray. Thus, this could be generalized and applied to other forms of entertainment.

Finally, our results in adaptivity provides a baseline reference for other studies aimed at investigating new techniques to improve classification performance, such as wavelet analysis (Nicolas-Alonso and Gomez-Gil, 2012) and RNN (Krusienski et al., 2011). Alternatively, researches in game adaptation could investigate the fusion between physiological signals based adaptation and user performance based adaptation for better performance. Reproducing our method for other serious gaming could also provide an illustration of the adaptation effect towards the player's exposure to educational content and learning. Hierarchical adaptivity mechanisms can be investigated to provide fine-grained difficulty ranges by taking account different complexity of gaming systems, game interfaces, and other design factors (Ahn et al., 2014).

Appendices

Appendix A

The flow questionnaires

Table A.1: GameFlow questionnaire (adapted from Kiili (2006)).

No	Item	Indicator
1	I knew clearly what I wanted to do and achieve.	Clear goal
2	The goals of the game were clearly defined.	
3	I was challenged, but I believed my skills would allow me to meet the challenge.	Challenge/task
4	The challenge that the game provided and my skills were at an equally high level.	
5	I could use the game user interface spontaneously and automatically without having to think.	Playability/artefact
6	The game user interface was ease to use.	
7	I could tell by the way I was performing how well I was doing.	Immediate feedback
8	I was aware how I was performing in the game.	
9	I felt in total control of my actions.	Sense of control
10	I had a feeling of control of my actions.	
11	My attention was focused entirely on playing the game.	Concentration
12	It was no effort to keep my mind on game events.	
13	I had total concentration while playing the game.	
14	I was not concerned with what others may have been thinking about my playing performance.	Loses of self-awareness
15	I was not worried about my performance during playing.	
16	I was totally immersed in playing the game.	
17	I really enjoyed the playing experience.	Rewarding experience
18	I loved the feeling of playing and want to capture it again.	

Continued on next page

Table A.1 – continued from previous page

No	Item	Indicator
19	The playing experience left me feeling great.	
20	I found the experience extremely rewarding.	
21	The way time passed seemed to be different from normal.	Time distortion
22	My sense of time altered (either speeded up or slowed down).	
23	I experienced a clear flow during playing.	Overall flow
24	Open ended questions: • If you experienced flow, what factors in the game contributed to flow? • If you did not experience flow, what factors in the game disturbed achieving flow?	

Table A.2: EGameFlow questionnaire (adapted from Fu et al. (2009)).

Indicator	Id	Item
Concentration	C1	*The game captures my attention*[d]
	C2	*The game content captures my attention*[d]
	C3	Most of the gaming activities are related to the learning task
	C4	No distraction from the task is highlighted
	C5	In general, I remain concentrated in the game
	C6	I am not distracted from the tasks that I should concentrate on
	C7	I am not burdened with tasks that seem unrelated
	C8	Workload in the game is adequate
Goal Clarity	G1	Overall goals were presented in the beginning of the game
	G2	Overall goals were presented clearly
	G3	Intermediate goals were presented in the beginning of each scene
	G4	Intermediate goals were presented clearly
	G5	*I understand the learning goal through the game*[d]
Feedback	F1	I receive feedback on my progress in the game
	F2	I receive immediate feedback on my actions

Continued on next page

APPENDIX A: THE FLOW QUESTIONNAIRES

Table A.2 – continued from previous page

Indicator	Id	Item
	F3	I am notified of new tasks immediately
	F4	I am notified of new events immediately
	F5	I receive information on my status (success/failure) of intermediate goals immediately
	F6	*I receive information on my status, such as score or level*[d]
Challenge	H1	*I enjoy the game without feeling bored or anxious*[d]
	H2	*The challenge is adequate, neither too easy nor too difficult*[d]
	H3	The game provides text hints to help me overcome the challenges
	H4	The game provides "online support" that help me overcome the challenges
	H5	The game provides video/audio auxiliaries to help me overcome the challenges
	H6	*My skill improves through the course of overcoming the challenges*[d]
	H7	*I am driven by my skills improvement*[d]
	H8	The difficulty of challenges increase as my skills improved
	H9	The game provides new challenges with proper pacing
	H10	The game provides different levels of challenges that tailor to different players
Autonomy	A1	*I feel a sense of control over the menu*[d]
	A2	*I feel a sense of control over actions of roles or objects*[d]
	A3	*I feel a sense of control over interactions between roles or objects*[d]
	A4	*The game does not allow players to make errors to a degree such that they cannot progress in the game*[d]
	A5	*The game supports my recovery from errors*[d]
	A6	*I feel that I can use my strategies freely*[d]
	A7	I feel a sense of control and impact over the game
	A8	I know next step in the game
	A9	I feel a sense of control over the game
Immersion	I1	I forget about time passing while playing the game
	I2	I become unaware of my surroundings while playing the game

Continued on next page

Table A.2 – continued from previous page

Indicator	Id	Item
	I3	I temporary forget worries about everyday life while playing the game
	I4	I experience an altered sense of time
	I5	I become involved in the game
	I6	I feel emotionally involved in the game
	I7	I feel viscerally involved in the game
Social interaction	S1	I feel cooperative toward other players
	S2	I collaborate intensively with other players
	S3	The cooperation in the game is helpful for learning
	S4	The game supports social interaction between players
	S5	The game supports communities within the game
	S6	The game supports communities outside the game
Knowledge improvement	K1	The game increases my knowledge
	K2	I get the basic ideas of the knowledge taught
	K3	I try to apply the knowledge in the game
	K4	The game motivates the player to integrate the knowledge taught
	K5	I want to know more about the knowledge taught

[d] *item was deleted after reliability test*

The learning outcomes test set

Table B.1: Question items for assessing conceptual knowledge

Principle	Id	Item
Newton's first law	1	The book remains at rest on the table because ... (a) of the gravity (b) the net force is equal to zero (c) the table prevents the book from falling down
	2	A car is moving on a straight road with a constant velocity. The sum of force acting on it ... (a) has the same direction with the car (b) depends on the car speed (c) is zero (d) is equal to the weight of the car (e) depends on the mass of the car
	3	When a car is moving with a constant velocity to the right, the total force acting on it ... (a) is equal to zero (b) has direction to the right (c) has direction to the left

Continued on next page

Table B.1 – continued from previous page

Principle	Id	Item
Newton's second law	4	A man applies a constant horizontal force on the empty and non moving cart. Rain starts falling vertically into it. Which one is correctly described the motion of the cart until it is filled with water (assuming any friction force is absent) ... (a) the cart's acceleration is constant (b) the cart's acceleration is continuously decreasing (c) the cart is moving with a constant velocity
	5	Two identical boxes are lying on smooth horizontal surfaces, one on earth and the other on the moon. We want to give both boxes the same horizontal acceleration. The required force is (assuming any friction force is absent) ... (a) the same for the two bodies (b) bigger on the earth (c) bigger on the moon
Newton's third law	6	A pot is lying on a table and acting a force on the table with downwards direction. The reaction to this force is ... (a) the force from the earth to the pot (b) the force from the table to the pot (c) the weight of the pot to the earth

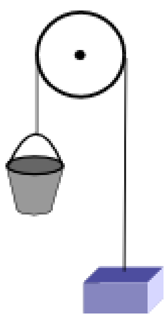
Continued on next page

Table B.1 – continued from previous page

Principle	Id	Item
	7	A box is hanged to the roof with a rope. The reaction to the weight of the box (which has downward direction) is the force acting from ... (a) the box to the rope (b) the roof to the rope (c) the rope to the box (d) the box to the earth
	8	While you are standing on a balance with your shoes on, you pull the laces of your shoes, the indication of the balance will ... (a) become smaller (b) become bigger (c) remain the same
The concept of force	9	A golf ball is moving in the air after being hit. Students claim that there are three forces acting on the ball: the gravity (B), the force of the knock (F), and the force of the air resistance (T). However, the force on the ball is in fact the sum of ... (a) B only (b) B and F (c) B and T (d) F and T (e) all of them

Continued on next page

Table B.1 – continued from previous page

Principle	Id	Item
Mass vs. weight	10	A stone is weighted on the surface of the earth and the moon.In which place the stone has bigger weight? (a) the earth (b) the moon (c) the same in both places
	11	A stone was weighted on the surface of the earth and another stone was weighted on the moon. Both showed the same weights. Which stone has larger mass? (a) the stone on the earth (b) the stone on the moon (c) the same for both stones
	12	The bucket and the box in the figure below have the same masses but are staying at different heights above the floor. Which one has the bigger weight? (a) the box (b) the bucket (c) the same 

Continued on next page

Table B.1 – continued from previous page

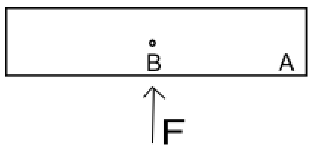
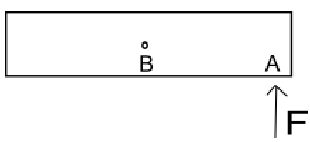
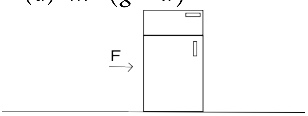
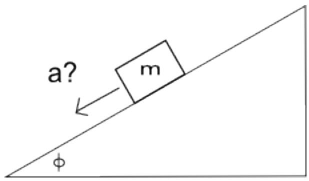
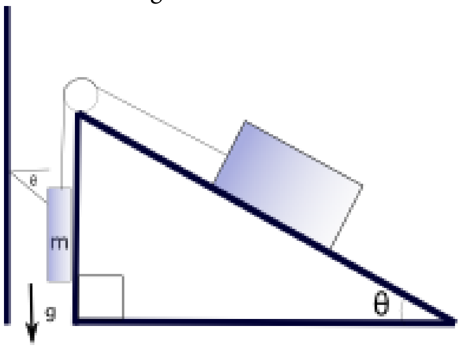
Principle	Id	Item
Force and torque	13	What will happen if you apply a perpendicular force F to the center of the mass (B) of the eraser on the table (assuming uniform density)? (a) it will shift (b) it will rotate (c) it will shift and rotate (d) nothing will happen 
	14	What will happen if you apply a perpendicular force F to the point A of the eraser on the table (assuming uniform density)? (a) it will shift (b) it will rotate (c) it will shift and rotate (d) nothing will happen 

Table B.2: Question items for assessing procedural knowledge

Id	Item
1	Let g be the magnitude of the acceleration due to gravity, m is the mass of the refrigerator. Assuming there is no friction, what is the magnitude of the force F (in Newton) to move the refrigerator with acceleration of a? (a) $m \cdot g$ (b) $m \cdot a$ (c) $m \cdot (g + a)$ (d) $m \cdot (g - a)$ 
2	After the refrigerator starts to move to the left, what magnitude of the force (in Newton) to keep it moving in a constant velocity? (assuming there is no friction) (a) $m \cdot a$ (b) zero (c) $m \cdot (g + a)$ (d) $m \cdot (g - a)$

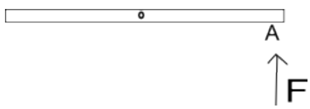
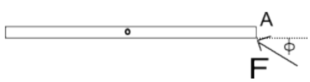
Continued on next page

Table B.2 – continued from previous page

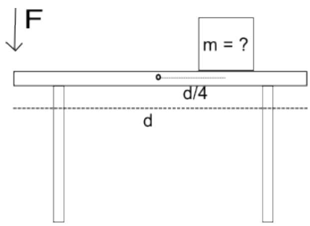
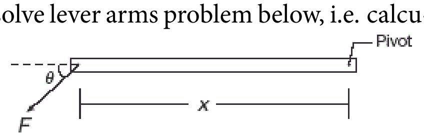
Id	Item
3	Let g be the magnitude of the acceleration due to gravity. a block with mass of m kg slides down on φ degree inclined plane. The kinetic friction between the floor and the box is μ_k. What is the acceleration of the box in m/s^2? (a) $\mu_k \cdot g \cdot sin\varphi$ (b) $\mu_k \cdot g \cdot cos\varphi$ (c) $g \cdot (sin\varphi - \mu_k \cdot cos\varphi)$ (d) $g \cdot (cos\varphi - \mu_k \cdot sin\varphi)$ 
4	What steps you will take to solve balanced force problem below, i.e. calculating the net force of the box on the right hand side? 

Continued on next page

Table B.2 – continued from previous page

Id	Item
5	Assume the stick has length of d, negligible width, and uniform density (pivot in the center of mass). If we apply a force perpendicular F to the point A, what is the torque value (in Nm)? (a) $F \cdot d$ (b) $F \cdot \frac{d}{2}$ (c) $F \cdot \frac{d}{4}$ (d) $F \cdot 2d$ 
6	Assume the eraser has length of d with uniform density (pivot in the center of mass). If we apply a force F to the point A, what is the torque value (in Nm)? (a) $F \cdot d \cdot sin\varphi$ (b) $F \cdot \frac{d}{2} \cdot sin\varphi$ (c) $F \cdot d \cdot sin(180 - \varphi)$ (d) $F \cdot \frac{d}{2} \cdot sin(180 - \varphi)$ 

Continued on next page

Id	Item
7	Let g be the magnitude of the acceleration due to gravity and d the length of the table. What is the appropriate mass m for the box to keep the table in balance given force F is applied at the opposite side of the table (assuming the table has uniform density)? (a) $F/2$ (b) $F/2g$ (c) $2F$ (d) $2F/g$ 
8	What steps you take to solve lever arms problem below, i.e. calcu-  lating the net torque?

Bibliography

Abdi, H. and Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459.

Adams, W., Reid, S., and LeMaster, R. (2008a). A Study of Educational Simulations Part I-Engagement and Learning. *Journal of Interactive Learning*, 19(3):397–419.

Adams, W., Reid, S., and LeMaster, R. (2008b). A study of educational simulations Part II-Interface Design. *Journal of Interactive Learning*, 19(4):551–577.

Adams, W. K., Paulson, A., Wieman, C. E., Henderson, C., Sabella, M., and Hsu, L. (2008c). What Levels of Guidance Promote Engaged Exploration with Interactive Simulations? In *AIP Conference Proceedings*, pages 59–62. Aip.

Ahn, R., Barakova, E., Feijs, L., Funk, M., Hu, J., and Rauterberg, M. (2014). Interfacing with adaptive systems. *Automation, Control and Intelligent Systems*, 2(4):53–61.

Aleven, V. A. and Koedinger, K. R. (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, 26(2):147–179.

Alfieri, L., Brooks, P. J., Aldrich, N. J., and Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, 103(1):1–18.

Amory, A. (2007). Game object model version ii: a theoretical framework for educational game development. *Educational Technology Research and Development*, 55(1):51–77.

Amory, A. and Seagram, R. (2003). Educational game models: conceptualization and evaluation: the practice of higher education. *South African Journal of Higher Education*, 17(2):p–206.

Ampatzoglou, A. and Chatzigeorgiou, A. (2007). Evaluation of object-oriented design patterns in game development. *Inf. Softw. Technol.*, 49(5):445–454.

Andersson, B. (1986). The experiential gestalt of causation: A common core to pupils' preconceptions in science. *European Journal of Science Education*, 8(2):155–171.

Andrews, S., Mora, J., Lang, J., and Lee, W. S. (2006). Hapticast: A physically-based 3d game with haptic feedback. *Proceedings of FuturePlay*.

Ang, K. K., Chin, Z. Y., Wang, C., Guan, C., and Zhang, H. (2012). Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b. *Frontiers in neuroscience*, 6.

Annetta, L. A. (2010). The "i's" have it: A framework for serious educational game design. *Review of General Psychology*, 14(2):105.

Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., and Christopherson, R. (2009). Emotion sensors go to school. In *AIED*, volume 200, pages 17–24.

Babiloni, F., Cincotti, F., Mattia, D., De Vico Fallani, F., Tocci, A., Bianchi, L., Salinari, S., Marciani, M., Colosimo, A., and Astolfi, L. (2007). High resolution eeg hyperscanning during a card game. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 4957–4960. IEEE.

Beale, I. L., Kato, P. M., Marin-Bowling, V. M., Guthrie, N., and Cole, S. W. (2007). Improvement in cancer-related knowledge following use of a psychoeducational video game for adolescents and young adults with cancer. *Journal of Adolescent Health*, 41(3):263–270.

Belliveau, J., Kennedy, D., McKinstry, R., Buchbinder, B., Weisskoff, R., Cohen, M., Vevea, J., Brady, T., and Rosen, B. (1991). Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*, 254(5032):716–719.

Bellotti, F., Berta, R., De Gloria, A., D'ursi, A., and Fiore, V. (2012). A serious game model for cultural heritage. *Journal on Computing and Cultural Heritage (JOCCH)*, 5(4):17.

Bellotti, F., Berta, R., De Gloria, A., and Primavera, L. (2009a). Adaptive experience engine for serious games. *Computational Intelligence and AI in Games, IEEE Transactions On*, 1(4):264–280.

Bellotti, F., Berta, R., De Gloria, A., and Primavera, L. (2009b). A task annotation model for sandbox serious games. In *Computational Intelligence and Games, 2009. CIG 2009. IEEE Symposium on*, pages 233–240.

Bellotti, F., Berta, R., De Gloria, A., and Primavera, L. (2010). Supporting authors in the development of task-based learning in serious virtual worlds. *British Journal of Educational Technology*, 41(1):86–107.

Berger, H. (1929). Über das elektrenkephalogramm des menschen. *European Archives of Psychiatry and Clinical Neuroscience*, 87(1):527–570.

Berta, R., Bellotti, F., Gloria, A. D., Pranantha, D., and Schatten, C. (2013). Electroencephalogram and physiological signal analysis for assessing flow in games. *IEEE Trans. Comput. Intellig. and AI in Games*, 5(2):164–175.

Bianchi-Berthouze, N., Kim, W. W., and Patel, D. (2007). Does body movement engage you more in digital game play? and why? In *Affective Computing and Intelligent Interaction*, pages 102–113. Springer.

Blanchard, G., Bousquet, O., and Massart, P. (2008). Statistical performance of support vector machines. *The Annals of Statistics*, pages 489–531.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Muller, K.-R. (2008). Optimizing spatial filters for robust eeg single-trial analysis. *Signal Processing Magazine, IEEE*, 25(1):41–56.

Bloor, R., Baroudi, C., Kaufman, M., et al. (2007). *Service oriented architecture for dummies*. John Wiley & Sons.

Blume, W. T., Holloway, G. M., Kaibara, M., and Young, G. B. (2010). *Blume's Atlas of pediatric and adult electroencephalography*. Lippincott Williams and Wilkins.

Blunt, R. (2007). Does game-based learning work? results from three recent studies. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*. NTSA.

Bonestroo, W. J. and de Jong, T. (2012). Effects of planning on task load, knowledge, and tool preference: a comparison of two tools. *Interactive Learning Environments*, 20(2):141–153.

Bos, D. P.-O., Reuderink, B., Van de Laar, B., Gürkök, H., Mühl, C., Poel, M., Nijholt, A., and Heylen, D. (2010). Brain-computer interfacing and games. In *Brain-Computer Interfaces*, pages 149–178. Springer.

Boucsein, W. (2012). *Electrodermal activity*. Springer.

Brawner, K. W. and Goldberg, B. S. (2012). Real-time monitoring of ecg and gsr signals during computer-based training. In *Proceedings of the 11th International Conference on Intelligent Tutoring Systems*, ITS'12, pages 72–77, Berlin, Heidelberg. Springer-Verlag.

Broin, D. O. (2011). Using a criteria-based user model for facilitating flow in serious games. In *Games and Virtual Worlds for Serious Applications (VS-GAMES), 2011 Third International Conference on*, pages 63–69. IEEE.

Brown, D. E. (1989). Students' concept of force: the importance of understanding newton's third law. *Physics Education*, 24(6):353.

Cameron, B. and Dwyer, F. (2005). The effect of online gaming, cognition and feedback type in facilitating delayed achievement of different learning objectives. *Journal of Interactive Learning Research*, 16(3):243–258.

Carlson, N. R. (2012). *Physiology of Behavior 11th Edition*. Pearson.

Chanel, G., Kronegg, J., Grandjean, D., and Pun, T. (2006). Emotion assessment: Arousal evaluation using eeg's and peripheral physiological signals. In *Multimedia content representation, classification and security*, pages 530–537. Springer.

Chanel, G., Rebetez, C., Bétrancourt, M., and Pun, T. (2011). Emotion assessment from physiological signals for adaptation of game difficulty. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 41(6):1052–1063.

Chang, C.-C. and Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.

Chaouachi, M., Jraidi, I., and Frasson, C. (2011). Modeling mental workload using eeg features for intelligent systems. In Konstan, J., Conejo, R., Marzo, J., and Oliver, N., editors, *User Modeling, Adaption and Personalization*, volume 6787 of *Lecture Notes in Computer Science*, pages 50–61. Springer Berlin Heidelberg.

Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM*, 50(4):31–34.

Cohen, D. (1968). Magnetoencephalography: evidence of magnetic fields produced by alpha-rhythm currents. *Science*, 161(3843):784–786.

Connolly, T. M., Boyle, E. a., MacArthur, E., Hainey, T., and Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2):661–686.

Coyle, D., Garcia, J., Satti, A. R., and McGinnity, T. M. (2011). Eeg-based continuous control of a game using a 3 channel motor imagery bci: Bci game. In *Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2011 IEEE Symposium on*, pages 1–7. IEEE.

Csikszentmihalyi, I. S. (1992). *Optimal experience: Psychological studies of flow in consciousness*. Cambridge University Press.

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., and Lang, P. J. (2000). Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biological psychology*, 52(2):95–111.

De Freitas, S., Jameson, J., et al. (2012). *The e-learning reader*. The Continuum International Publishing Group/Bloomsbury Publishing Plc.

De Freitas, S. and Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education*, 46(3):249–264.

De Jong, T. (2006). Technological Advances in Inquiry Learning. *Science (New York, N.Y.)*, 312(5773):532–3.

Delorme, A. and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21.

Derbali, L., Chalfoun, P., and Frasson, C. (2011). A theoretical and empirical approach in assessing motivational factors: From serious games to an its. In *International Florida Artificial Intelligence Research Society Conference*, pages 513–518.

Deuschl, G. and Eisen, A. (1999). Recommendations for the practice of clinical neurophysiology(guidelines of the international federation of clinical neurophysiology). *Electroencephalography and clinical neurophysiology. Supplement*.

Dimberg, U. (1990). Facial electromyography and emotional reactions. *Psychophysiology*.

D'mello, S. and Graesser, A. (2013). Autotutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Trans. Interact. Intell. Syst.*, 2(4):1–39.

Einthoven, W. (1902). Galvanometrische registratie van het menschelijk electrocardiogram. *Herinneringsbundel Professor SS Rosenstein*, pages 101–107.

Elam, K. (2004). *Grid Systems: Principles of Organizing Type (Design Briefs)*. Princeton Architectural Press, New York.

Fatourechi, M., Bashashati, A., Ward, R. K., and Birch, G. E. (2007). {EMG} and {EOG} artifacts in brain computer interface systems: A survey. *Clinical Neurophysiology*, 118(3):480 – 494.

Feeney, E. J. (2007). Quality feedback: The essential ingredient for teacher success. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 80(4):191–198.

Fenz, W. and Epstein, S. (1967). Changes in gradients of skin conductance, heart rate and respiration rate as a function of experience. *Psychosomatic Medicine*, 29:33–51.

Ferguson-hessler, M. G. M. and de Jong, T. (1991). A Model of the cognitive aspects of physics instruction. In *AERA Annual Meeting*, pages 3–7.

Finneran, C. M. and Zhang, P. (2003). A person–artefact–task (pat) model of flow antecedents in computer-mediated environments. *International Journal of Human-Computer Studies*, 59(4):475–496.

Fisch, B. J. and Spehlmann, R. (1999). *Fisch and Spehlmann's EEG primer: basic principles of digital and analog EEG*. Elsevier Health Sciences.

Friedman, N., Geiger, D., and Goldszmidt, M. (1997). Bayesian network classifiers. *Machine learning*, 29(2-3):131–163.

Fu, F.-L., Su, R.-C., and Yu, S.-C. (2009). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education*, 52(1):101–112.

Galili, I. (1995). Mechanics background influences students' conceptions in electromagnetism. *International Journal of Science Education*, 17(3):371–387.

Garay, N., Cearreta, I., López, J. M., and Fajardo, I. (2006). Assistive technology and affective mediation. *Human Technology: an Interdisciplinary Journal on Humans in ICT Environments*, 2(1):55–83.

Garrett, D., Peterson, D. A., Anderson, C. W., and Thaut, M. H. (2003). Comparison of linear, nonlinear, and feature selection methods for eeg signal classification. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 11(2):141–144.

Garris, R., Ahlers, R., and Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & gaming*, 33(4):441–467.

Gijlers, H. and de Jong, T. (2013). Using Concept Maps to Facilitate Collaborative Simulation-Based Inquiry Learning. *Journal of the Learning Sciences*, 22(3):340–374.

Girouard, A., Solovey, E. T., Hirshfield, L. M., Chauncey, K., Sassaroli, A., Fantini, S., and Jacob, R. J. (2009). Distinguishing difficulty levels with non-invasive brain activity measurements. In *Human-Computer Interaction–INTERACT 2009*, pages 440–452. Springer.

Göbel, S., Hardy, S., Wendel, V., Mehm, F., and Steinmetz, R. (2010). Serious games for health: personalized exergames. In *Proceedings of the international conference on Multimedia*, pages 1663–1666. ACM.

Gouizi, K., Bereksi Reguig, F., and Maaoui, C. (2011). Emotion recognition from physiological signals. *Journal of medical engineering & technology*, 35(6-7):300–307.

Graesser, A., Chipman, P., Leeming, F., and Biedenbach, S. (2009). Deep learning and emotion in serious games. *Serious games: Mechanisms and effects*, pages 83–102.

Green, C. S. and Bavelier, D. (2006). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, 101(1):217–245.

Gunter, G. A., Kenny, R. F., and Vick, E. H. (2006). A case for a formal design paradigm for serious games. *International Digital Media and Arts Association*, 3(1):93–105.

Gürkök, H. and Nijholt, A. (2012). Brain-computer interfaces for multimodal interaction: A survey and principles. *International Journal of Human-Computer Interaction*, 28(5):292–307.

Haas, L. (2003). Hans berger (1873–1941), richard caton (1842–1926), and electroencephalography. *Journal of Neurology, Neurosurgery & Psychiatry*, 74(1):9–9.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., and Lounasmaa, O. V. (1993). Magnetoencephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics*, 65(2):413.

Heise, D. (2004). Enculturating agents with expressive role behavior. In *In Trappl, Robert. Agent Culture: Human-Agent Interaction in a Multicultural World*, pages 127–142. Lawrence Erlbaum Associates.

Hestenes, D. and Wells, M. (1992). A mechanics baseline test. *The physics teacher*, 30(3):159–166.

Hestenes, D., Wells, M., and Swackhamer, G. (1992). Force concept inventory. *The physics teacher*, 30(3):141–158.

Heyworth, R. (1999). International Journal of Procedural and conceptual knowledge of expert and novice students for the solving of a basic problem in chemistry. *International Journal of Science Education*, 21(2):195–211.

Hornik, K., Meyer, D., and Karatzoglou, A. (2006). Support vector machines in r. *Journal of statistical software*, 15(9):1–28.

Hsu, C.-W., Chang, C.-C., and Lin, C.-J. (2003). A practical guide to support vector classification. Technical report, Department of Computer Science and Information Engineering, National Taiwan University.

Hunicke, R., LeBlanc, M., and Zubek, R. (2004). Mda: A formal approach to game design and game research. In *Proceedings of the AAAI Workshop on Challenges in Game AI*, pages 04–04.

Ivory, J. D. and Kalyanaraman, S. (2007). The effects of technological advancement and violent content in video games on players' feelings of presence, involvement, physiological arousal, and aggression. *Journal of Communication*, 57(3):532–555.

Janicke, S. and Ellis, A. (2011). Psychological and physiological differences between the 3d and 2d gaming experience. In *Proceedings of the 3D Entertainment Summit*.

Jasczak, R. (1988). Tomographic radiopharmaceutical imaging. *Proceedings of the IEEE*, 76(9):1079–1094.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A. (2008). Measuring and defining the experience of immersion in games. *International journal of human-computer studies*, 66(9):641–661.

Jiang, L. and Zhang, H. (2006). Weightily averaged one-dependence estimators. In *PRICAI 2006: trends in artificial intelligence*, pages 970–974. Springer.

Jiang, L., Zhang, H., and Cai, Z. (2009). A novel bayes model: hidden naive bayes. *Knowledge and Data Engineering, IEEE Transactions on*, 21(10):1361–1371.

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Computer Assisted Learning*, 7(2):75–83.

Joyce, C. A., Gorodnitsky, I. F., and Kutas, M. (2004). Automatic removal of eye movement and blink artifacts from eeg data using blind component separation. *Psychophysiology*, 41(2):313–325.

Karpicke, J. D. and Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science (New York, N.Y.)*, 331(6018):772–5.

Kelly, H., Howell, K., Glinert, E., Holding, L., Swain, C., Burrowbridge, A., and Roper, M. (2007). How to build serious games. *Communications of the ACM*, 50(7):44–49.

Kiili, K. (2005a). Content creation challenges and flow experience in educational games: The it-emperor case. *Internet and Higher Education*, 8(3):183–198.

Kiili, K. (2005b). On educational game design: Building blocks of flow experience. *Tampereen teknillinen yliopisto. Julkaisu-Tampere University of Technology. Publication; 571.*

Kiili, K. (2006). Evaluations of an experiential gaming model. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments*, 2(2):187–201.

Kiili, K., de Freitas, S., Arnab, S., and Lainema, T. (2012). The design principles for flow experience in educational games. *Procedia Computer Science*, 15(0):78 – 91. 4th International Conference on Games and Virtual Worlds for Serious Applications(VS-GAMES'12).

Kiili, K. and Lainema, T. (2008). Foundation for measuring engagement in educational games. *Journal of Interactive Learning Research*, 19(3):469–488.

Kim, K. H., Bang, S., and Kim, S. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical and biological engineering and computing*, 42(3):419–427.

Kirriemuir, J. and McFarlane, A. (2004). Literature review in games and learning. *FutureLab Series*.

Kirschner, P. and Clark, R. (2006). Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, 41(2):75–86.

Kivikangas, J. M., Chanel, G., Cowley, B., Ekman, I., Salminen, M., Järvelä, S., and Ravaja, N. (2011). A review of the use of psychophysiological methods in game research. *Journal of Gaming & Virtual Worlds*, 3(3):181–199.

Kligfield, P., Gettes, L. S., Bailey, J. J., Childers, R., Deal, B. J., Hancock, E. W., Van Herpen, G., Kors, J. A., Macfarlane, P., Mirvis, D. M., et al. (2007). Recommendations for the standardization and interpretation of the electrocardiogrampart i: The electrocardiogram and its technology a scientific statement from the american heart association electrocardiography and arrhythmias committee, council on clinical cardiology; the american college of cardiology foundation; and the heart rhythm society endorsed by the international society for computerized electrocardiology. *Journal of the American College of Cardiology*, 49(10):1109–1127.

Kline, S. and Arlidge, A. (2003). Online gaming as emergent social media: A survey. *Technical report, Media Analysis Laboratory*.

Kolb, D. A. et al. (1984). *Experiential learning: Experience as the source of learning and development*, volume 1. Prentice-Hall Englewood Cliffs, NJ.

Kolloffel, B. and de Jong, T. (2013). Conceptual understanding of electrical circuits in secondary vocational engineering education: Combining traditional instruction with inquiry learning in a virtual. *Journal of Engineering Education*, 102(3):375 – 393.

Koster, R. (2013). *Theory of fun for game design*. O'Reilly Media, Inc.

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into practice*, 41(4):212–218.

Krusienski, D. J., Grosse-Wentrup, M., Galán, F., Coyle, D., Miller, K. J., Forney, E., and Anderson, C. W. (2011). Critical issues in state-of-the-art brain–computer interface signal processing. *Journal of neural engineering*, 8(2):025002.

Lajoie, S. P. (2005). Extending the scaffolding metaphor. *Instructional Science*, 33(5-6):541–557.

Lanzetta, J. T., Cartwright-Smith, J., and Eleck, R. E. (1976). Effects of nonverbal dissimulation on emotional experience and autonomic arousal. *Journal of personality and social psychology*, 33(3):354.

Larkin, J., McDermott, J., Simon, D. P., and Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, 208(4450):1335–1342.

Lazzaro, N. (2004). Why we play games: Four keys to more emotion without story. *Technical report*.

Levin, R. B. (2000). Devices and methods for maintaining an alert state of consciousness through brain wave monitoring. US Patent 6,167,298.

Li, S. (2010). Mind reading is on the market. *Los Angeles Times*.

Liu, C., Agrawal, P., Sarkar, N., and Chen, S. (2009). Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *International Journal of Human-Computer Interaction*, 25(6):506–529.

Lopes, R. and Bidarra, R. (2011). Adaptivity challenges in games and simulations: a survey. *Computational Intelligence and AI in Games, IEEE Transactions on*, 3(2):85–99.

Lucas, K. and Sherry, J. L. (2004). Sex differences in video game play: A communication-based explanation. *Communication Research*, 31(5):499–523.

Malone, T. (1981). What makes computer games fun? *SIGSOC Bull.*, 13(2-3):143–.

Mandryk, R. and Atkins, M. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies*, 65(4):329–347.

Mandryk, R. L., Atkins, M. S., and Inkpen, K. M. (2006a). A continuous and objective evaluation of emotional experience with interactive play environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 1027–1036, New York, NY, USA. ACM.

Mandryk, R. L. and Inkpen, K. M. (2004). Physiological indicators for the evaluation of co-located collaborative play. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, CSCW '04, pages 102–111, New York, NY, USA. ACM.

Mandryk, R. L., Inkpen, K. M., and Calvert, T. W. (2006b). Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & IT*, 25(2):141–158.

Manning, C. D., Raghavan, P., and Schutze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, pages 31–48.

Mayo, M. J. (2007). Games for science and engineering education. *Communications of the ACM*, 50(7):30–35.

McLaren, B., Lim, S., and Koedinger, K. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.). In Love, B., editor, *30th Annual Conference of the Cognitive Science Society*, pages 2176–2181.

Meyer, D., Leisch, F., and Hornik, K. (2003). The support vector machine under test. *Neurocomputing*, 55(1-2):169–186. Support Vector Machines.

Moreno, R. and Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19(3):309–326.

Nacke, L. E., Stellmach, S., and Lindley, C. A. (2011). Electroencephalographic assessment of player experience a pilot study in affective ludology. *Simulation & Gaming*, 42(5):632–655.

Nicolas-Alonso, L. F. and Gomez-Gil, J. (2012). Brain computer interfaces, a review. *Sensors*, 12(2):1211–1279.

Niedermeyer, E. and da Silva, F. L. (2004). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, 5th edition.

Nijboer, F., Allison, B., Dunne, S., Plass-Oude Bos, D., Nijholt, A., and Haselager, P. (2011). A preliminary survey on the perception of marketability of brain-computer interfaces and initial development of a repository of bci companies. In *5th International Brain-Computer Interface Conference, BCI 2011*. Verlag der Technischen Universitaet Graz.

Nijboer, J., Dorlas, J., and Mahieu, H. (1981). Photoelectric plethysmography-some fundamental aspects of the reflection and transmission methods. *Clinical Physics and Physiological Measurement*, 2(3):205.

Nte, S. and Stephens, R. (2008). Videogame aesthetics and e-learning: a retro-looking computer game to explain the normal distribution in statistics teaching. In *Proceedingsof the 2nd European conference on games-based learning (ECGBL)*, pages 16–17.

Obbink, M., Gürkök, H., Bos, D. P.-O., Hakvoort, G., Poel, M., and Nijholt, A. (2012). Social interaction in a cooperative brain-computer interface game. In *Intelligent Technologies for Interactive Entertainment*, pages 183–192. Springer.

Obikwelu, C., Read, J., and Sim, G. (2013). Children's problem-solving in serious games: The" fine-tuning system (fts)" elaborated. *Electronic Journal of e-Learning*, 11(1).

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, 2011:156869.

Orvis, K. A., Horn, D. B., and Belanich, J. (2007). Task difficulty and prior videogame experience: Their role in performance and motivation in instructional videogames. Technical report, DTIC Document.

Palatini, P. (1999). Need for a revision of the normal limits of resting heart rate. *Hypertension*, 33(2):622–625.

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education*, 52(1):1–12.

Paras, B. and Bizzocchi, J. (2005). Game, motivation, and effective learning: An integrated model for educational game design. In *Changing Views: Worlds in Play, 2005 International Conference*. DiGRA.

Peirce, N., Conlan, O., and Wade, V. (2008). Adaptive educational games: Providing non-invasive personalised learning experiences. In *Digital Games and Intelligent Toys Based Education, 2008 Second IEEE International Conference on*, pages 28–35. IEEE.

Pellouchoud, E., Smith, M. E., McEvoy, L., and Gevins, A. (1999). Mental effort-related eeg modulation during video-game play: Comparison between juvenile subjects with epilepsy and normal control subjects. *Epilepsia*, 40(s4):38–43.

Perkins, K., Adams, W., Dubson, M., Finkelstein, N., Reid, S., Wieman, C., and LeMaster, R. (2006). PhET: Interactive Simulations for Teaching and Learning Physics. *The Physics Teacher*, 44(1):18.

Pfurtscheller, G., Brunner, C., Schlögl, A., and Lopes da Silva, F. (2006). Mu rhythm (de) synchronization and eeg single-trial classification of different motor imagery tasks. *Neuroimage*, 31(1):153–159.

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures.(Trans A. Rosin).* Viking.

Piekarski, W. and Thomas, B. (2002). Arquake: the outdoor augmented reality gaming system. *Communications of the ACM*, 45(1):36–38.

Plotnikov, A., Stakheika, N., Schatten, C., Pranantha, D., Berta, R., and De Gloria, A. (2012). Measuring enjoyment in games through electroencephalogram (eeg) signal analysis. In *6th European Conference on Games-Based Learning (ECGBL 2012)*, pages 393–400.

Pranantha, D., Bellotti, F., Berta, R., and De Gloria, A. (2012a). Puzzle-it: An html5 serious games platform for education. In *E-Learning and Games for Training, Education, Health and Sports*, pages 134–143. Springer.

Pranantha, D., Bellotti, F., Berta, R., and DeGloria, A. (2012b). A Format of Serious Games for Higher Technology Education Topics: A Case Study in a Digital Electronic System Course. In *International Conference on Advanced Learning Technologies*, pages 13–17, Rome. IEEE.

Pranantha, D., Chen, W., Bellotti, F., Van der Spek, E., De Gloria, A., and Rauterberg, M. (2014). Designing physics game to support inquiry learning and to promote retrieval practice. In *6th International Conference on Computer Supported Education (CSEDU)*.

Prensky, M. (2002). The motivation of gameplay: The real twenty-first century learning revolution. *On the horizon*, 10(1):5–11.

Pulvermüller, F., Birbaumer, N., Lutzenberger, W., and Mohr, B. (1997). High-frequency brain activity: its possible role in attention, perception and language processing. *Progress in neurobiology*, 52(5):427–445.

Pun, T., Alecu, T. I., Chanel, G., Kronegg, J., and Voloshynovskiy, S. (2006). Brain-computer interaction research at the computer vision and multimedia laboratory, university of geneva. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 14(2):210–213.

Putman, P., Van Peer, J., Maimari, I., and Van der Werff, S. (2010). Eeg theta/beta ratio in relation to fear-modulated response-inhibition, attentional control, and affective traits. *Biological psychology*, 83(2):73–78.

Quinn, C. (1994). Designing educational computer games. in k. beattie, c. mcnaught and s. wills (eds). *Interactive multimedia in university education: Designing for change in teaching and learning*, pages 45–57.

Raîche, G., Walls, T. A., Magis, D., Riopel, M., and Blais, J.-G. (2013). Non-graphical solutions for cattell's scree test. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 9(1):23.

Rajaraman, A. and Ullman, J. D. (2011). *Mining of Massive Datasets*. Cambridge University Press.

Rakotomamonjy, A. and Guigue, V. (2008). Bci competition iii: dataset ii-ensemble of svms for bci p300 speller. *Biomedical Engineering, IEEE Transactions on*, 55(3):1147–1154.

Ratan, R. and Ritterfeld, U. (2009). Classifying serious games. *Serious games: Mechanisms and effects*, pages 10–24.

Ravaja, N., Turpeinen, M., Saari, T., Puttonen, S., and Keltikangas-Järvinen, L. (2008). The psychophysiology of james bond: Phasic emotional responses to violent video game events. *Emotion*, 8(1):114.

Rollings, A. and Adams, E. (2003). *Andrew Rollings and Ernest Adams on game design*. New Riders.

Rosenberg, L. B. and Riegel, J. R. (2002). Haptic feedback for touchpads and other touch controls. US Patent 6,429,846.

Salen, K. and Zimmerman, E. (2003). *Rules of Play: Game Design Fundamentals*. MIT Press, Cambridge, Massachusetts.

Salminen, M. and Ravaja, N. (2008). Increased oscillatory theta activation evoked by violent digital game events. *Neuroscience letters*, 435(1):69–72.

Savinainen, A., Scott, P., and Viiri, J. (2005). Using a bridging representation and social interactions to foster conceptual change: Designing and evaluating an instructional sequence for newton's third law. *Science Education*, 89(2):175–195.

Scheirer, J., Fernandez, R., Klein, J., and Picard, R. W. (2002). Frustrating the user on purpose: a step toward building an affective computer. *Interacting with computers*, 14(2):93–118.

Shapiro, M. A. and Peña, J. (2009). Generalizability and validity in digital game research. *Serious Games: Mechanisms and Effects*, page 389.

Sheikholeslami, C., Yuan, H., He, E., Bai, X., Yang, L., and He, B. (2007). A high resolution eeg study of dynamic brain activity during video game play. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 2489–2491.

Sheth, R. (2003). Avatar technology: giving a face to the e-learning interface. *The eLearning Developers' Journal*, pages 1–10.

Shim, B.-S., Lee, S.-W., and Shin, J.-H. (2007). Implementation of a 3-dimensional game for developing balanced brainwave. In *Software Engineering Research, Management & Applications, 2007. SERA 2007. 5th ACIS International Conference on*, pages 751–758. IEEE.

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, 78(1):153–189.

Sicart, M. (2008). Defining game mechanics. *Game Studies*, 8(2):1–14.

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology*, 64(2):489–528.

Sliney, A. and Murphy, D. (2008). Jdoc: A serious game for medical learning. In *Advances in Computer-Human Interaction, 2008 First International Conference on*, pages 131–136. IEEE.

Society, S. (2014). Serious games web services catalogue. http://services.seriousgamessociety.org/. Accessed: 2014-08-31.

Spronck, P., Sprinkhuizen-Kuyper, I., and Postma, E. (2004). Difficulty scaling of game ai. In *Proceedings of the 5th International Conference on Intelligent Games and Simulation (GAME-ON 2004)*, pages 33–37.

Squire, K., Barnett, M., Grant, J. M., and Higginbotham, T. (2004). Electromagnetism supercharged!: Learning physics with digital simulation games. In *Proceedings of the 6th international conference on Learning sciences*, pages 513–520. International Society of the Learning Sciences.

Steinmetz, H., Volkmann, J., Jäncke, L., and Freund, H.-J. (1991). Anatomical left-right asymmetry of language-related temporal cortex is different in left- and right-handers. *Annals of neurology*, 29(3):315–319.

Stylos, G., Evangelaki, G. A., and Kotsis, K. T. (2010). Misconceptions on classical mechanics by freshman university students: A case study in a physics department in greece. *Themes in Science and Technology Education*, 1(2):pp–157.

Sun, H., Xiang, Y., Sun, Y., Zhu, H., and Zeng, J. (2010a). On-line eeg classification for brain-computer interface based on csp and svm. In *Image and Signal Processing (CISP), 2010 3rd International Congress on*, volume 9, pages 4105–4108. IEEE.

Sun, H., Xiang, Y., Sun, Y., Zhu, H., and Zeng, J. (2010b). On-line eeg classification for brain-computer interface based on csp and svm. In *Image and Signal Processing (CISP), 2010 3rd International Congress on*, volume 9, pages 4105–4108.

Sward, K. A., Richardson, S., Kendrick, J., and Maloney, C. (2008). Use of a web-based game to teach pediatric content to medical students. *Ambulatory Pediatrics*, 8(6):354–359.

Sweetser, P. and Wyeth, P. (2005). Gameflow: A model for evaluating player enjoyment in games. *Comput. Entertain.*, 3(3):3–3.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285.

Sweller, J., Van Merrienboer, J. J., and Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, 10(3):251–296.

Tan, C. H., Tan, K. C., and Tay, A. (2011). Dynamic game difficulty scaling using adaptive behavior-based ai. *Computational Intelligence and AI in Games, IEEE Transactions on*, 3(4):289–301.

Tanzer, I. O. et al. (2006). *Numerical modeling in electro-and magnetoencephalography*. PhD thesis, Helsinki University of Technology.

Tao, J. and Tan, T. (2005). Affective computing: A review. In *Affective computing and intelligent interaction*, pages 981–995. Springer.

Terry, C. and Jones, G. (1986). Alternative frameworks: Newton's third law and conceptual change. *European Journal of Science Education*, 8(3):291–298.

Thornton, R. K. and Sokoloff, D. R. (1998). Assessing student learning of newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics*, 66(4):338–352.

Treado, P. J., Levin, I. W., and Lewis, E. N. (1992). Near-infrared acousto-optic filtered spectroscopic microscopy: a solid-state approach to chemical imaging. *Applied spectroscopy*, 46(4):553–559.

Van de Laar, B., Gurkok, H., Plass-Oude Bos, D., Poel, M., and Nijholt, A. (2013). Experiencing bci control in a popular computer game. *Computational Intelligence and AI in Games, IEEE Transactions on*, 5(2):176–184.

Van der Spek, E. D. (2011). *Experiments in serious game design: a cognitive approach*. PhD thesis, Utrecht University.

Van der Spek, E. D. (2012). Towards designing for competence and engagement in serious games. In *Serious Games Development and Applications*, pages 98–109. Springer.

Van der Spek, E. D., Wouters, P., and Van Oostendorp, H. (2011). Code red: Triage or cognition-based design rules enhancing decisionmaking training in a game environment. *British Journal of Educational Technology*, 42(3):441–455.

Van Erp, J. B., Lotte, F., Tangermann, M., et al. (2012). Brain-computer interfaces: beyond medical applications. *Computer-IEEE Computer Society-*, 45(4):26–34.

Van Lankveld, G. (2013). *Quantifying Individual Player Differences*. PhD thesis, Tilburg University.

Van Lankveld, G., Spronck, P., and Rauterberg, M. (2008). Difficulty scaling through incongruity. In *the Fourth Artificial Intelligence and Interactive Digital Entertainment Conference*, pages 228–229. AAAI Press.

Van Lankveld, G., Spronck, P., Van Den Herik, H. J., and Rauterberg, M. (2010). Incongruity-based adaptive game balancing. In *Advances in computer games 2009*, pages 208–220. Springer.

Van Lehn, K., Lynch, C., and Schulze, K. (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence in Education*, 15(3):147–204.

Van Oostendorp, H., Van der Spek, E., and Linssen, J. (2014). Adapting the complexity level of a serious game to the proficiency of players. *EAI Endorsed Transactions on Serious Games*, 14(2).

Virvou, M., Katsionis, G., and Manos, K. (2005). Combining software games with education: Evaluation of its educational effectiveness. *Journal of Educational Technology & Society*, 8(2).

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., and Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, 34(3):229–243.

Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., and Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and instruction*, 11(4):381–419.

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard university press.

Wang, Q., Sourina, O., and Nguyen, M. K. (2010). Eeg-based" serious" games design for medical applications. In *Cyberworlds (CW), 2010 International Conference on*, pages 270–276. IEEE.

Weibel, D., Wissmath, B., Habegger, S., Steiner, Y., and Groner, R. (2008). Playing online games against computer-vs. human-controlled opponents: Effects on presence, flow, and enjoyment. *Computers in Human Behavior*, 24(5):2274–2291.

Wensveen, S. A. G., Djajadiningrat, J. P., and Overbeeke, C. J. (2004). Interaction frogger: a design framework to couple action and function through feedback and feedforward. In *Designing interactive systems: processes, practices, methods, and techniques*, pages 177–184, New York, NY, USA. ACM.

Wilson, G. M. and Sasse, M. A. (2000a). Do users always know what's good for them? utilising physiological responses to assess media quality. In *Proc. HCI 2000: People and Computer XIV - Usability or else!*, pages 327–339. Springer.

Wilson, G. M. and Sasse, M. A. (2000b). Investigating the impact of audio degradations on users: Subjective vs. objective assessment methods. In *Proc. OZCHI 2000*, pages 135–142.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clinical neurophysiology*, 113(6):767–791.

Wood, J. (2011). *Communication in our lives*. Cengage Learning.

Woolf, B. P. (2009). *Building Intelligent Interactive Tutors for revolutionizing e-learning*. Morgan Kaufmann, Amsterdam.

Woolf, B. P., Arroyo, I., Muldner, K., Burleson, W., Cooper, D. G., Dolan, R., and Christopherson, R. M. (2010). The effect of motivational learning companions on low achieving students and students with disabilities. In *Intelligent Tutoring Systems*, pages 327–337. Springer.

Wouters, P., Van der Spek, E., and Van Oostendorp, H. (2009). Current practices in serious game research: a review from a learning outcomes perspective. *Games-based learning advancements for multi-sensory human computer interfaces: techniques and effective practices*, pages 232–250.

Wouters, P., Van Nimwegen, C., Van Oostendorp, H., and Van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 105(2):249–265.

Yaman, M., Nerdel, C., and Bayrhuber, H. (2008). The effects of instructional support and learner interests when learning using computer simulations. *Computers & Education*, 51(4):1784–1794.

Yannakakis, G. and Hallam, J. (2007). Capturing player enjoyment in computer games. In Baba, N., Jain, L., and Handa, H., editors, *Advanced Intelligent Paradigms in Computer Games*, volume 71 of *Studies in Computational Intelligence*, pages 175–201. Springer Berlin Heidelberg.

Yun, C., Shastri, D., Pavlidis, I., and Deng, Z. (2009). O' game, can you feel my frustration?: Improving user's gaming experience via stresscam. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 2195–2204, New York, NY, USA. ACM.

Zimmerman, J., Thiene, P., and Harding, J. (2003). Design and operation of stable rf-biased superconducting point-contact quantum devices, and a note on the properties of perfectly clean metal contacts. *Journal of Applied Physics*, 41(4):1572–1580.

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer*, 38(9):25–32.

Zyda, M., Mayberry, A., Wardynski, C., Shilling, R., and Davis, M. (2003). The moves institute's america's army operations game. In *Proceedings of the 2003 symposium on Interactive 3D graphics*, pages 219–220. ACM.

Summary

In the field of serious games (SG), there is a clear need for supporting pedagogical authors with methodologies and tools that can support them in providing effective learning experiences. Exploring this challenge, there was a number of successful SG that tends to provide players with suited knowledge structures for investigating a specific educational domain. These SG were defined through an abstract model for facilitating authors in creating adaptive contents. Based on this, Serious Games Society (SGS) is initiating the creation of services to support efficient serious games development under a Service Oriented Architecture (SOA). The goal is to provide serious game developers with a repository of a well documented and ready-to-use services (either SOAP or RESTful) usable to develop serious games following the SOA paradigm. This will prevent researchers and developers from reinventing the wheel since certain functionalities for their game may have already existed as services. Our work focused on two key aspects of efficient serious games development: the format and the content delivery strategy.

Firstly, a game format was designed and developed under an SOA platform, and the flow framework was used to define the game features implemented in the platform. A game prototype was then developed with a specific, clear, and quantifiable learning goal and evaluated in term of user perception and performance. This was necessary to improve the prototype for the subsequent phase, i.e. evaluation of game features in terms of flow and learning. Subsequently, using the game prototype we altered a game feature (or service) and measure its effect on flow and learning. This is essential to evaluate the usefulness of game features in terms of flow and learning for either game design improvement or game services selection. Consequently, tools are needed to evaluate flow and learning from playing the games. To this end, we reviewed several questionnaires for evaluating flow and devised a test set to quantify learning outcomes. Using the game prototype and the assessment tools, we measured the effect of a game feature, i.e. a tutoring tool, on both flow and learning in games. This was performed by comparing two game prototypes for learning Physics: with a tutor and without a tutor. Our hypothesis was that the one with the tutor would have lower flow since it is likely to obstruct player in proceeding with the game. We found that the two game prototypes have significantly different flow in which, surprisingly, the one with the tutor has higher flow. Furthermore, we found flow improves the perceived performance, but not the actual

performance. Consequently, we argue that having a sense of better learning with higher flow does not necessarily contribute to real learning.

Secondly, the content delivery strategy in games is important to provide balance between the level of challenges and the player's skills. However, designing a balanced game becomes highly complex as the size of the potential audience grows since different players have different skills and they expect different challenges. In this case, adaptivity mechanisms become necessary to regulate the delivery of challenges. Player and task modeling could be useful for representing the player's ability and the difficulty of challenges for adaptivity, but they are still contingent to the characteristics of the audiences. On the other hand, physiological signals may serve as an alternative or provide additional information for adaptivity since human share similar physiological traits in many circumstances. Thus, we performed experiments on the use of physiological signals to support adaptivity in games.

To this end, we collected physiological signals (brainwaves, heart rate, and skin conductance) of several users while playing games in three different settings: boredom, flow, and frustration. Subsequently, we trained two types of classifiers: a) collective classifier for classifying the state of all players, and b) individual classifier for classifying the state of each player. We found that flow can be distinguished from boredom and frustration using 1-s window of brainwave signals at moderate level. This implies the possibility of real time inference of player state in consumer context given the time for data collection and classifier training time, and the real time difficulty adaptation.

About the author

Danu Pranantha Dolar was born and raised in Ponorogo, Jawa Timur, Indonesia. Upon finishing high school in 1999, he continued his study in the Department of Informatics Engineering at Bandung Institute of Technology (ITB). He completed his bachelor degree in 2003 (*cum laude*). Subsequently, he worked at Honda Motor as an IT analyst from 2004 to 2008. During his working period at Honda from 2005 to 2007, he received a graduate scholarship from Korea Institute of Science and Technology (KAIST) where he obtained his master degree in multimedia computing, communication, and broadcasting. He authored 2 conference papers during his master program. In 2009, Danu joined Institut Teknologi Sepuluh Nopember Surabaya (ITS), as a lecturer in the Department of Information System. There, he taught a variety of courses including algorithm and programming, database, intelligent system, data mining, and IT service management.

In January 2011, Danu started his doctorate study under Erasmus Mundus Joint Doctorate Program (EMJD) in Interactive and Cognitive Environment (ICE). His primary university was University of Genoa (UNIGE), Italy and his secondary university was Eindhoven University of Technology (TU/e), the Netherlands. During his doctorate, Danu was involved in the EU Games and Learning Alliance (GALA) project organized by UNIGE, where he worked on services for serious games (SG) development. This led him to authoring this dissertation.

Publications

Submitted or to be submitted

1. Pranantha, D., Van der Spek, E. D., Bellotti, F., Berta, R., De Gloria, A., and Rauterberg, M. Extensible physics based simulation game with a tutoring system for learning Newton's principles. to be submitted to Computational Intelligence and AI in Games, IEEE Transaction on, special issue Physics-Based Simulation Games.

2. Pranantha, D., Van der Spek, E. D., Bellotti, F., Berta, R., De Gloria, A., and Rauterberg, M. Evaluating game features on the flow and learning. to be submitted to Entertainment Computing.

3. Pranantha, D., Berta, R., Bellotti, F., De Gloria, A., Van der Spek, E. D., and Rauterberg, M. (2014). Using brainwaves for real-time content adaptation in games. submitted to SPIE newsroom.

Published or accepted for publication

1. Pranantha, D., Luo, C., Bellotti, F., de Gloria, A. (2011). Designing Contents for a Serious Game for Learning Computer Programming with Different Target Users. In 2011th International Conference on Design and Modeling in Science, Education, and Technology (DeMSET).
2. Luo, C., Espinosa, AP., Pranantha, D., De Gloria, A. (2011). Multi-robot search and rescue team. Safety, Security, and Rescue Robotics (SSRR), IEEE International Symposium on , pages 296-301.
3. Pranantha, D., Bellotti, F., Berta, R., and DeGloria, A. (2012). A Format of Serious Games for Higher Technology Education Topics: A Case Study in a Digital Electronic System Course. In International Conference on Advanced Learning Technologies, pages 13-17, Rome. IEEE.
4. Pranantha, D., Bellotti, F., Berta, R., and De Gloria, A. (2012). Puzzle-it: An html5 serious games platform for education. In E-Learning and Games for Training, Education, Health and Sports, pages 134-143. Springer.
5. Plotnikov, A., Stakheika, N., Schatten, C., Pranantha, D., Berta, R., and De Gloria, A. (2012). Measuring enjoyment in games through electroencephalogram (eeg) signal analysis. In 6th European Conference on Games-Based Learning (ECGBL 2012), pages 393-400.
6. Berta, R., Bellotti, F., De Gloria, A., Pranantha, D., Schatten, C. (2013). Electroencephalogram and physiological signal analysis for assessing flow in games. Computational Intelligence and AI in Games, IEEE Transactions on, 5(2), 164-175.
7. Ling. Y., Pranantha, D. (2013). HTML5 Serious Games Platform for Education, IEEE Technology and Engineering Education (ITEE), 8(2), pages 15-17.
8. Pranantha, D., Chen, W., Bellotti, F., Spek, E. v., De Gloria, A., and Rauterberg, M. (2014). Designing physics game to support inquiry learning and to promote retrieval practice. In 6th International Conference on Computer Supported Education (CSEDU).
9. Pranantha, D., Spek, E. v., Bellotti, F., Berta, R., De Gloria, A., and Rauterberg, M. (2014). Game Design and Development for Learning Physics Using the Flow Framework. In 2014 Gaming and Learning Alliance (GALA) Conference.

Other published publications

1. Pranantha, D., Kim, M., Hahm, S., Kim, B., Lee, K., Park, K. (2007). Dependent quantization for scalable video coding. In Advanced Communication Technology, The 9th International Conference on, 222-227. IEEE.

2. Mahendrawathi, E. R., Pranantha, D., Utomo, J. D. (2010). Development of dashboard for hospital logistics management. In Open Systems (ICOS), 2010 IEEE Conference on, 86-90. IEEE.